



Università degli Studi di Pisa

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea Magistrale in Conservazione ed Evoluzione

TESI DI LAUREA

**Mesotelina: regolazione post-trascrizionale e analisi
dell'evoluzione del gene.**

Candidato:

Giulia Pasquesi Irene Maria

Matricola 439930

Relatori:

Dott. Stefano Landi

Dott. Sergio Tofanelli

*Può darsi che non ci siano delle risposte.
A tutti coloro che ancora credono, e cercano.
Ad ogni predatore che insegue la sua preda.*

Abstract

La mesotelina è una proteina ancorata alla membrana cellulare, caratterizzata da un profilo di espressione tessuto-specifica nel mesotelio. Tuttavia, questa è presente anche in tessuti tumorali di origine non mesoteliale (ovaio e pancreas) a livelli superiori di quelli fisiologici, come dimostra il confronto in contesto tessuto-specifico tra pleure sane ed affette da mesotelioma pleurico maligno (MPM). Non è nota la funzione biologica da questa svolta, ed al momento le maggiori indicazioni sono di un possibile coinvolgimento in processi di adesione cellulare eterotipica, proliferazione ed invasività.

Se si confrontano i livelli di messaggero tra tessuti e linee cellulari di mesotelio sano e di MPM si osserva un rapporto non proporzionale con i corrispettivi livelli di espressione proteica, che è molto inferiore nelle linee sane. Si è ipotizzata pertanto una regolazione ad opera di microRNA, presenti a livello fisiologico e sottoespressi nel tumore.

La ricerca bibliografica su lavori high-throughput di microRNA profiling ha portato alla selezione del microRNA 204-5p quale possibile candidato della regolazione post-trascrizionale della mesotelina.

La capacità regolatoria del microRNA è stata testata in vitro in cellule della linea mero-14 secondo una metodica di real-time RT-PCR per i livelli di espressione del messaggero e di western blot per la proteina.

Al fine di rendere più solide le predizioni sulla funzionalità regolatoria, l'analisi sperimentale è stata integrata da una analisi in-silico sul microRNA e sul cDNA del gene *MSLN*, volte a valutare il grado di conservazione evolutiva dell'interazione microRNA-messaggero. È stata quindi condotta una analisi esone per esone sulla sequenza della mesotelina in 17 specie di mammiferi al fine di identificare accelerazioni evolutive per siti e taxa specifici.

Dall'analisi è emersa una significativa capacità del microRNA di reprimere l'espressione del gene tanto a livello di trascritto quanto di proteina, con siti di nucleazione predetti nella regione codificante; il fatto che la sequenza del microRNA maturo sia poi altamente conservata nei vertebrati supporta il suo valore funzionale. La porzione di trascritto che codifica per la mesotelina matura presenta solo pochi siti sottoposti ad accelerazione evolutiva nei mammiferi, due per l'esone 14 ed uno per l'esone 16, e sembra che nel complesso sia piuttosto sottoposta a selezione purificante. Apparentemente, in tempi recenti la mesotelina non ha acquisito nuove caratteristiche funzionali o strutturali, ma non si esclude che nuove caratteristiche possano essere insorte prima della divergenza dei mammiferi dagli altri vertebrati.

Sono stati infine identificati tre siti di nucleazione microRNA-mRNA perfettamente conservati nei mammiferi, ad ulteriore supporto del ruolo regolatorio del microRNA 204-5p nei confronti della mesotelina.

Abstract

Mesothelin is a cell surface glycoprotein expressed specifically by mesothelial cells, but also at a greater extent by cancer non-mesothelial tissues (ovary and pancreas), as can be stated comparing healthy pleurae and malignant pleural mesothelioma (MPM) samples. Although the biological role has not been validated yet, there are clues that mesothelin may participate in heterotypic cellular adhesion, proliferation and invasiveness.

The difference of the messenger's expression level between tissues and cell lines of healthy mesothelium and MPM does not correlate with the protein concentration, which is far lower in healthy samples. The explanation has been implied to a negative regulation by microRNA in normal conditions, and disregulated in tumors.

Referring to high-throughput microRNA profiling papers, the microRNA 204-5p has been selected as a possible candidate for mesothelin's post-transcriptional regulation, since it resulted down-regulated in MPM, ovarian and pancreatic carcinomas.

The effectiveness of the microRNA has been tested in vitro in cells of the mero-14 cell line, following a real-time RT-PCR protocol for the mRNA, and western blot for the protein.

To support the predicted functionality, the experimental work has been strengthened by an in-silico analysis both for the microRNA and the cDNA of the *mesothelin* gene, aimed to evaluate the degree of evolutionary conservation for the microRNA-mRNA interaction.

It has been chosen to perform an exon-by-exon analysis of the mesothelin sequence in 17 mammalian species, so to identify positive selection for specific sites and taxa.

A significant negative regulation exerted by microRNA 204-5p has been found both at the messenger and protein levels, with nucleation sites predicted in the coding region of the mesothelin mRNA.

The high conservation of the mature microRNA sequence in almost all vertebrate lineages suggests its regulative role to be relevant.

The region encoding for the mature mesothelin peptide shows just few codons subjected to positive selection in Mammalia, two in exon 14 and one in exon 16; it could therefore be ruled out that it rather experienced a strong purifying selection. Although it seems that mesothelin did not acquire structural or functional discriminative features in late times, it is not ruled out that this could have happened before the mammalian divergence from the other vertebrate lineages.

Three nucleation sites highly conserved between the placentals' species have been identified, strongly supporting the regulatory role exerted by the microRNA 204-5p on mesothelin.

Ringraziamenti

Sono passati due anni, e ancora vivo la medesima circostanza.

No, non la stessa; la laurea triennale rappresentava un momento di passaggio, un primo piccolo ma cruciale traguardo. Se allora portavo con me ricordi di cambiamenti, e di brillanti aspettative, oggi la sensazione è differente. Sento che si chiude una porta.

Per l'immediato, mi si prospetta un ritorno a casa, e allo stesso momento un andar via da casa. Lascio il mondo che ho in qualche modo plasmato, e che mi ha senza dubbio modellata negli ultimi 5 anni. Vorrei che fosse un arrivederci, piuttostochè un addio. Perché qui rimangono compagni ed amici con i quali ho condiviso momenti ed occasioni importanti. Qui, rimane una famiglia. E no, non voglio che sia un addio.

Ma sono dei ringraziamenti questi, non le mie memorie, quindi per una volta proverò ad attenermi all'argomento.

Per primo, ringrazio il mio personale rasoio di Occam; senza, l'apprendimento e la scoperta, gli esami.. tutto sarebbe stato meno avvincente. Certo, forse la mia carriera sarebbe stata un poco più lineare, più facile, ma l'incentivo allora? No, non recrimino questo mio modo di essere.

Chi ha dovuto confrontarsi però sono i professori, e fosse solo per questo, meritano un riconoscimento speciale. Ringrazio il professor Landi ed il professor Tofanelli, primi tra tutti, perchè è con loro che ho avuto il piacere e l'onore di relazionarmi più strettamente, come studentessa a lezione e come allieva in laboratorio. Sono grata per tutto ciò che avete fatto per me; per la passione e per l'ispirazione. Spero che anche loro, come me, siano soddisfatti dell'esperimento portato avanti fino a questo punto. E ringrazio il professor Gabrielli, per la partecipazione alle fasi preliminari di studio, e per tutti i consigli in corso d'opera fino a questa stesura finale.

Grazie anche a tutti coloro che mi hanno aiutata in laboratorio, ad Elisa Bracci ed Ombretta, a Monica, a Maria Sole, ad Elisa.. senza i vostri consigli ed il supporto, non sarebbe stata una esperienza altrettanto bella, ed istruttiva.

Grazie ai professori Barbanera e Sbrana, a tutti coloro che hanno rafforzato la mia passione nella genetica, nella conservazione e nella zoologia; spero di aver fatto tesoro dei vostri insegnamenti, perchè anche da questi dipenderà chi diverrò. Anche se non lo sapranno mai, io potrò ricordarlo: la paleontologia è una nobile disciplina, ma oltre al fascino, non pensavo potesse coinvolgermi tanto, e riuscirci; professor Landini, e professor Ragaini, il merito a voi.

Perseguirò per quanto mi sarà possibile nel proposito che avevo due anni fa, quando ho iniziato: seguirò la passione. É questa la mia motivazione, il mio istinto; finché non ci sbatterò contro, il sogno sarà il primo pensiero.

Non è sufficiente ringraziare, devo riconoscervi il merito, il dono meraviglioso che forse inconsciamente mi avete affidato: il viaggio e la scoperta. Non avere paura, ma affrontare ciò che può essere; vivere, solo vivere. Mamma, Babbo, questo è per voi.

Questo però è “solo” il risultato della combinazione di tanti, piccoli ma basilari tasselli (questo insegna la genetica, questa è la biologia), tutti coloro che hanno fatto parte della vostra e della mia vita. Ai vostri genitori –i miei nonni e le mie nonne– alle vostre sorelle, alle mie cugine ed ai loro figli, ed a tutti i vostri amici che sono ora i miei “zii”, così cari che erano al mio fianco anche due anni fa.

Alla mia famiglia.

Mi avete sempre insegnato che il sangue è però solo una componente, allora ci sono altre famiglie, ideali o reali, che meritano un riconoscimento.

Alla Mickie e a Brambo; tra l'altro, come vi considero, reale o ideale? Reale, assolutamente.

Alla Luna, a Byron, alla mia Princ Lung Dao e a tutte le Princ che ancora lottano; se non mi avete lasciato segni indelebili voi...

Alla Betta, a Fede, a Luca, a Fabio, a Miglio, a Filo.. ho un ritorno importante in vista, siete stati il passato, vi ho perso nel recente, vorrei ritrovarvi nel futuro.

Federica, Vroni, Flavia, Filippo, Michele, Luca. Per quanto vi conosca da poco, e di sicuro non abbastanza, mi avete raggiunta, o vi siete lasciati avvicinare, in ogni caso siete stati i compagni di viaggio di questi ultimi anni, e vi ringrazio, dal profondo del cuore.

S, non so neanche cosa potrei dire a te, o di te.. Quanto più ci sia in me che non sia io. Mario, Terry, Marghe, nonna Dorian, nonna Giovanna, Mau.. mi vanto del privilegio di poter pensare a voi come alla mia seconda famiglia, reale ed ideale. Non avrei potuto sperare di intrecciare la mia vita con persone più meravigliose. Grazie.

E tra i compagni, vi sono Fabio e Lorenzo: il vostro amico Giulia vi spalleggerà sempre.

Un ringraziamento va a quella bambina ribelle che su una spiaggia si lasciava guidare dall'istinto e dall'incoscienza, incamminandosi da sola con la sola compagnia di un cane; a quella ragazza che ha scelto i propri viaggi sulla base della passione; a quella quasi-donna che ha deciso di perseverare, comunque. A questa me stessa che ancora si meraviglia e si commuove.

Ed infine, ma solo perché davvero non ho modo di pensare quale altro merito attribuirvi se non la "conclusione ad effetto", a Chiara. Sei stata il capo, la compagna, l'ispirazione, la forza, la gioia, il rigore, l'impegno, la passione ed il caposaldo di tutto. Ti voglio bene, e grazie, grazie mille. Sappiamo tutte che la mia aspirazione è altrove, ma se rifarei tutto da capo senza alcuna differenza, il merito è solo tuo.

"And I will reignite..."

Indice

Ringraziamenti	ix
1 Introduzione	1
1.1 La Mesotelina	2
1.1.1 Origine e relazioni evolutive	2
1.1.2 Dal gene al trascritto primario, fino alla proteina	3
1.1.3 Promotore e regolazione della trascrizione	7
1.2 Genomica e genetica comparata	9
1.2.1 La genomica comparata	9
1.2.2 Applicazioni della genomica e della genetica comparata	10
1.2.3 Omega	12
1.2.4 La mesotelina umana ed i suoi ortologi	14
1.3 I microRNA	16
1.3.1 Biogenesi dei microRNA	16
1.3.2 Principi di interazione microRNA-mRNA	18
1.3.3 Modalità di azione	19
1.3.4 Considerazioni evolutive	24
1.3.5 Il microRNA miR204-5p	25
2 Materiali e metodi	29
2.1 Il microRNA miR-204-5p	30

2.1.1	Selezione del microRNA da testare	30
2.1.2	Linea cellulare mero-14	30
2.1.3	Protocollo sperimentale	31
2.1.4	Analisi del messaggero	32
2.1.5	Analisi della proteina	34
2.2	Analisi comparata in-silico del gene <i>MSLN</i>	37
2.2.1	Allineamenti	37
2.2.2	Generazione degli alberi filogenetici	38
2.2.3	Analisi filogenetica ed evolutiva	40
2.3	Conservazione filogenetica del miR-204	43
3	Risultati	45
3.1	Il microRNA miR-204-5p	45
3.1.1	Espressione del messaggero	45
3.1.2	Analisi dei livelli di espressione proteica	48
3.2	Analisi comparata in-silico del gene <i>MSLN</i>	49
3.2.1	Analisi filogenetica ed evolutiva	50
3.3	Conservazione filogenetica del miR-204	55
4	Discussione	57
4.1	Considerazioni funzionali	57
4.2	Il microRNA 204-5p	61
4.2.1	Considerazioni evolutive sui microRNA	62
4.3	Analisi filogenetica	63
4.4	Cinetica di espressione	65
4.5	Considerazioni conclusive	66
5	Conclusioni	69
A	Transcrip Variant MSLN	71
B	Materiali e metodi microRNA	75
B.1	Selezione del microRNA da testare	75
B.2	Linea cellulare mero-14	77

B.3	Protocollo sperimentale	78
B.3.1	Semina (1° giorno)	78
B.3.2	Trasfezione (2° giorno)	79
B.3.3	Estrazione delle 24h (3° giorno)	80
B.4	Valutazione della qualità e della quantità dell'RNA	83
B.4.1	Analisi dei livelli di espressione del messaggero	83
B.4.2	Analisi dei livelli di espressione della proteina	89
B.5	Protocollo di Western Blot per 2 gel	92
B.5.1	Assemblaggio	92
B.5.2	Preparazione dei gel	93
B.5.3	Preparazione dei campioni	94
B.5.4	Caricamento dei campioni	95
B.5.5	Trasferimento su membrana	96
B.5.6	Trattamento della membrana	97
B.5.7	Applicazione dell'anticorpo primario	98
B.5.8	Applicazione dell'anticorpo secondario	98
B.5.9	Lettura al ChemiDoc MP System (Bio-Rad)	99
C	Materiali e metodi filogenesi	101
C.1	Scelta delle specie	101
C.2	Allineamenti	102
C.2.1	Generazione degli alberi filogenetici	104
C.3	Analisi filogenetica ed evolutiva	110
C.3.1	Preparazione dei file di input	111
C.3.2	Analisi Site Model	113
C.3.3	Analisi del significato delle sostituzioni aminoacidiche	114
C.3.4	Stima del valore Omega per il gene <i>MSLN</i>	116
D	Risultati microRNA	117
D.1	Selezione del microRNA 204-5p e prove preliminari	117
D.1.1	Selezione del microRNA	117
D.2	Quantificazione RNA miRNA-204-5p	120
D.3	Analisi statistica espressione dell'mRNA	123

D.4	Analisi dei livelli di espressione proteica	123
D.4.1	Quantificazione proteica WB mod1	123
E	Risultati analisi filogenetica	127
E.1	Generazione degli allineamenti	127
E.1.1	Allineamenti proteici	131
E.2	Alberi filogenetici	140
F	Risultati analisi comparata miRNA	149
F.0.1	Blocco1	149
F.0.2	Blocco2	151
G	Allineamenti MSLN	155
G.1	Allineamenti proteici	155
G.2	Allineamenti cDNA	166
H	Calcolo LRT	183
H.0.1	LRT M2a-M1a	184
H.0.2	LRT M8-M7	185
H.0.3	LRT M8-M8a	185
H.0.4	LRT M3-M0	186
I	Alberi filogenetici	187
I.1	Analisi della sequenza aminoacidica; E1–16	187
I.2	Analisi della sequenza del cDNA; E1–16	191
J	Analisi statistica completa	195
J.1	Analisi statistica tempo per tempo	195
J.1.1	ANOVA 2 fattori, 24h	195
J.1.2	ANOVA 2 fattori, 48h	197
J.1.3	ANOVA 2 fattori, 72h	199
J.2	Analisi statistica esperimento per esperimento	201
J.2.1	ANOVA 2 fattori, Esperimento 1	201
J.2.2	ANOVA 2 fattori, Esperimento 2	203
J.2.3	ANOVA 2 fattori, Esperimento 3	204

<i>INDICE</i>	xvii
J.2.4 ANOVA 2 fattori, Esperimento 4	206
J.2.5 ANOVA 2 fattori, Esperimento 5	207
Bibliografia	209

Elenco delle figure

1.1	Maturazione del precursore MSLN-MPF	3
1.2	Isoforme della mesotelina	6
1.3	Sviluppo nel tempo del numero di progetti genomici	9
1.4	Maturazione dei micro-RNA	18
1.5	Regolazione post-inizio traduzione dei microRNA	20
1.6	Regolazione dell'inizio traduzione ad opera dei microRNA	21
1.7	Attivazione trascrizionale ad opera dei microRNA	22
1.8	Degradazione del messaggero ad opera dei microRNA	23
1.9	Regolazione post-trascrizionale negativa ad opera dei microRNA	24
1.10	Il microRNA miR-204-5p	26
1.11	Allineamento miR-204-5p	28
3.1	Interaction plot mRNA mesotelina	46
3.2	Western Bolt, modello 1	48
3.3	Western Bolt, modello 2	49
3.4	Stato ancestrale per l'aa 482	53
4.1	Embrione di topo, 14.5dpc;MSLN	58
4.2	Embrione di topo, 14.5dpc;MSLNL	59
4.3	Livelli espressione MSLN	60
A.1	Isoforme NCBI	71

A.2	Isoforme Ensembl	73
A.3	Isoforme UCSC	74
B.1	jMiRanda.jar	77
B.2	Esempio di piastra p60	79
B.3	Separazione di fasi	81
B.4	Curva tipica di amplificazione di real time PCR	87
B.5	Assemblaggio dei supporti di Western Blot	93
B.6	Alloggiamento dei vetrini nel supporto di corsa	95
D.1	microRNA a comune leganti la 3'UTR	118
D.2	microRNA a comune, down-regolati	118
D.3	siti di nucleazione miR-204–MSLN	119
D.4	RNA estratti da cellule	120
D.5	Retta di taratura	124
E.1	Albero filogenetico; proteina, senza outgroup	141
E.2	Albero filogenetico; proteina, con outgroup	142
E.3	Albero filogenetico; cDNA, senza outgroup	143
E.4	Albero filogenetico; cDNA, con outgroup	144
E.5	Albero filogenetico; proteina 9–16, senza outgroup	145
E.6	Albero filogenetico; proteina 9–16, con outgroup	146
E.7	Albero filogenetico; cDNA 9–16, senza outgroup	147
E.8	Albero filogenetico, cDNA 9–16, con outgroup	148
I.1	Albero B1, proteina, senza outgroup	187
I.2	Albero B1, proteina, con outgroup	188
I.3	Albero B2, proteina, senza outgroup	189
I.4	Albero B2, proteina, con outgroup	190
I.5	Albero B1, cDNA, senza outgroup	191
I.6	Albero B1, cDNA, con outgroup	192
I.7	Albero B2, cDNA, senza outgroup	192
I.8	Albero B2, cDNA, con outgroup	193
J.1	Interaction plot 24h	196

J.2	Interaction plot 48h	197
J.3	Interaction plot 72h	199
J.4	Interaction plot Exp1	201
J.5	Interaction plot Exp2	203
J.6	Interaction plot Exp3	204
J.7	Interaction plot Exp4	206
J.8	Interaction plot Exp5	207

Elenco delle tabelle

1.1	Caratteristiche del gene nelle specie scelte per l'analisi.	15
3.1	Analisi della varianza	47
3.2	Analisi della varianza, tempo	47
3.3	Analisi PAML, tutte le specie	51
3.4	Analisi PAML, senza Opossum	51
3.5	Analisi PAML, senza Chimpanzee	52
3.6	Valori medi di omega	54
B.1	Protocollo di real time qPCR.	88
(a)	Protocollo di qPCR	88
(b)	Profilo di termociclizzazione	88
D.1	Estrazioni RNA exp1–2	121
D.2	Estrazioni RNA exp3–5	122
D.3	Least square means	123
D.4	Assorbanza standard	124
D.5	Quantificazione proteica modello 1	124
D.6	Quantificazione proteica modello 2	125
H.1	LRT M2a-M1a	184
(a)	LRT M2a-M1a con outgroup	184

	(b) LRT M2a-M1a senza outgroup	184
	(c) LRT M2a-M1a senza chimpanzee	184
H.2	LRT M8-M7	185
	(a) LRT M8-M7 con outgroup	185
	(b) LRT M8-M7 senza outgroup	185
	(c) LRT M7-M8 senza chimpanzee	185
H.3	LRT M8-M8a	185
	(a) LRT M8-M8a con outgroup	185
	(b) LRT M8-M8a senza outgroup	185
	(c) LRT M8-M8a senza chimpanzee	185
H.4	LRT M3-M0	186
	(a) LRT M3-M0 con outgroup	186
	(b) LRT M3-M0 senza outgroup	186
	(c) LRT M3-M0 senza chimpanzee	186
J.1	Analisi varianza 24h	196
J.2	Least squares means, 24h	197
J.3	Analisi varianza 48h	198
J.4	Least squares means, 48h	199
J.5	Analisi varianza 72h	200
J.6	Least squares means, 72h	200
J.7	Analisi varianza Exp1	201
J.8	Least squares means, Exp1	202
J.9	Analisi varianza Exp2	203
J.10	Least squares means, Exp2	204
J.11	Analisi varianza Exp3	205
J.12	Least squares means, Exp3	205
J.13	Analisi varianza Exp4	206
J.14	Least squares means, Exp4	207
J.15	Analisi varianza Exp5	208
J.16	Least squares means, Exp5	208

Introduzione

Indice

1.1	La Mesotelina	2
1.1.1	Origine e relazioni evolutive	2
1.1.2	Dal gene al trascritto primario, fino alla proteina	3
1.1.3	Promotore e regolazione della trascrizione	7
1.2	Genomica e genetica comparata	9
1.2.1	La genomica comparata	9
1.2.2	Applicazioni della genomica e della genetica comparata	10
1.2.3	Omega	12
1.2.4	La mesotelina umana ed i suoi ortologi	14
1.3	I microRNA	16
1.3.1	Biogenesi dei microRNA	16
1.3.2	Principi di interazione microRNA-mRNA	18
1.3.3	Modalità di azione	19
	Legame alla regione codificante	24
1.3.4	Considerazioni evolutive	24
1.3.5	Il microRNA miR204-5p	25

1.1 La Mesotelina

1.1.1 Origine e relazioni evolutive

Il gene *Mesotelina* ha avuto origine da un evento di duplicazione di un gene ancestrale in una fase precoce dell'evoluzione del subphylum *Vertebrata*, relazione supportata dal ritrovamento già nei rettili (*Anolis carolinensis* e *Pelodiscus sinensis*) di entrambi gli elementi derivati: i geni *MSLN* –Mesotelina– e *MSLNL* –Mesotelina-Like– (Kim & Hahn, 2011).

Nell'uomo i due geni sono localizzati con orientamento inverso sul cromosoma 16 (16p13.3), e presentano una organizzazione strutturale simile che potrebbe a sua volta sottendere ad una origine per duplicazione di un gene ancestrale comune (Daniels *et al.*, 2001).

Nel 2011 sono stati pubblicati i risultati di uno studio comparato per il gene *MSLNL* (Kim & Hahn, 2011), rilevando come in *Homo sapiens* non sia possibile identificare nè le sequenze del corrispettivo mRNA nè alcuna EST, presumibilmente a causa di rapida degradazione del trascritto o di silenziamento trascrizionale del gene stesso. Nella sequenza ricostruita mediante il confronto con la sequenza murina sono emerse diverse mutazioni deleterie nella ORF –open reading frame– umana, in particolare una transizione C>T nella tripletta CGA che porta all'introduzione di un codone di stop prematuro (TGA) nell'esone 10.

Sembrerebbe quindi che in *Homo sapiens* il gene, il cui trascritto è soggetto a nonsense-mediated mRNA decay, è in una fase evolutiva di degenerazione.

Non solo in *Homo sapiens*, ma anche in *Gorilla gorilla* sono state evidenziate diverse mutazioni deleterie per questo gene, che ne compromettono l'espressione.

Il rapporto dN/dS tra mutazioni non sinonime e mutazioni sinonime (riferirsi al paragrafo 1.2.3 a pagina 12) ricavato dall'analisi comparata degli orologi di uomo, chimpanzee, gorilla, macaca ed orangio ha suggerito una selezione rilassata sul gene *MSLNL* in queste linee, e che l'inattivazione potrebbe aver avuto luogo prima della divergenza tra le grandi antropomorfe africane.

1.1.2 Struttura della Mesotelina:

dal gene al trascritto primario, fino alla proteina

Il gene *mesotelina* (NC_000016.9) di *Homo sapiens* si compone di 17 esoni, di cui 16 codificanti: il primo e parte del secondo costituiscono la 5'UTR, con sito di inizio della traduzione in corrispondenza del nucleotide 168 del trascritto primario (RefSeq NM_005823, transcrip variant 1, mRNA). Il codone di stop, TGA, cade nell'esone 17, che include anche la 3'UTR.

Questo gene codifica per un precursore proteico di 620 amminoacidi (69KDa circa) composto di:

- una sequenza segnale per il reticolo endoplasmatico all'N-terminale di 36 amminoacidi ;
- una sequenza idrofobica al C-terminale di 24 amminoacidi che viene poi rimossa nella forma matura
- quattro potenziali siti di glicosilazione (amminoacidi 57; 388; 496; 523).

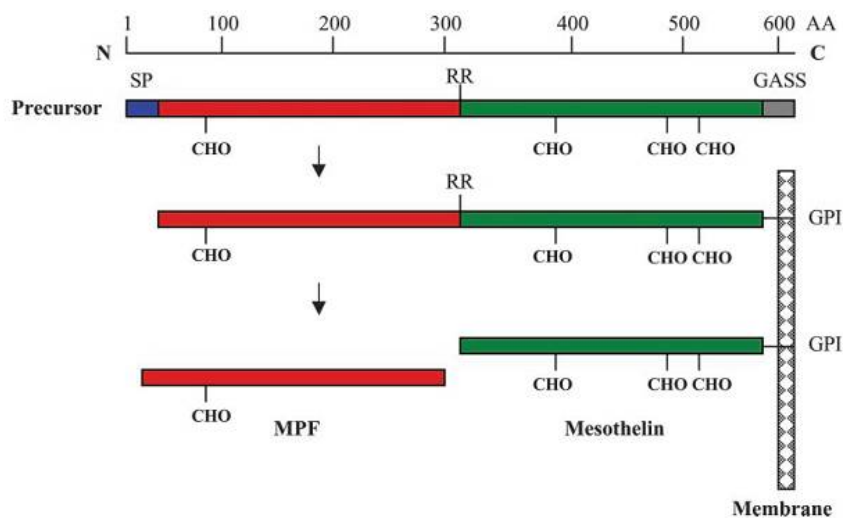


Figura 1.1: Rappresentazione schematica del processo di maturazione del precursore proteico, con il peptide segnale potenziale (SP) e la sequenza segnale dell'ancora GPI (GASS) predetti rispettivamente all'N-terminale ed al C-terminale.

Dopo che almeno uno dei siti potenziali è andato incontro a glicosilazione, il precursore è sottoposto all'azione di una proteina simil-furina che taglia tra gli amminoacidi 288–293 e determina la formazione di due frammenti, entrambi rappresentanti una proteina funzionale (illustrato in fig. 1.1 nella pagina precedente).

Il frammento all'N-terminale (amminoacidi 37–286) ha peso molecolare di 31KDa, ed è noto come fattore di potenziamento dei megacariociti (MPF). La proteina viene secreta nel sangue ed agisce come una citochina che ha la capacità di stimolare la formazione di colonie di megacariociti nel midollo osseo di topo in presenza dell'interleukina 3, ma non è noto se anche nell'uomo possa svolgere una funzione analoga (Yamaguchi *et al.*, 1994; Kojima *et al.*, 1995).

Il frammento al C-terminale (amminoacidi 296–606) rappresenta la mesotelina, a peso molecolare di 40KDa; a differenza dell'MPF, questa proteina è ancorata alla membrana cellulare per mezzo di una coda di glicosil-fosfatidil-inositolo (GPI) all'estremità carbossi-terminale.

La mesotelina mostra un'espressione tessuto-specifica limitata al mesotelio, la cui origine embrionale viene ricondotta al mesoderma tra l'ottavo ed il diciottesimo giorno di gestazione nell'uomo. Del mesotelio embrionale permangono nell'adulto le membrane sierose del pericardio, della pleura e del peritoneo, delimitanti le corrispettive cavità corporee.

La proteina è stata identificata inizialmente come target antigenico dell'anticorpo monoclonale MAb K1, e viene considerata come il marcatore di differenziamento delle cellule mesoteliali; nei saggi immunoistochimici tuttavia l'anticorpo non reagiva unicamente a livello di tessuto mesoteliale, ma anche di tessuti tumorali quali mesoteliomi e carcinoma ovarico (Chang & Pastan, 1996). Studi successivi hanno poi rivelato come a livello patologico l'espressione della mesotelina è maggiore di quanto non si riscontri nei tessuti mesoteliali, un dato ben evidente in pleure sane a confronto con quelle di mesotelioma pleurico maligno (MPM).

Poco è noto riguardo al ruolo biologico che la mesotelina svolge, ma sembra che nessuno dei due prodotti codificati sia essenziale, come suggerito da esperimenti di knock-out genico in topo dove non è stato riscontrato alcun fenotipo patologico, e tanto i maschi quanto le femmine si sviluppano e riproducono normalmente (Bera & Pastan, 2000).

Rump *et al.* (2004) hanno individuato il primo interattore proteico della mesotelina, la mucina MUC16, che è altamente espressa a livello di cellule tumorali ovariche. Considerando il contesto di espressione di entrambe, e che possono interagire così da mediare l'adesione eterotipica di cellule in vitro, è stato suggerito che l'interazione favorisca la metastatizzazione peritoneale del tumore. Nel 2009, Kaneko *et al.* (2009) hanno identificato una regione di 64 amminoacidi (296-359) all'N-terminale della mesotelina come il frammento minimo richiesto per la completa attività di legame di MUC16. A sostegno del possibile coinvolgimento nell'adesione cellulare, gli stessi autori hanno inoltre riscontrato come trattamenti con un anticorpo monoclonale specifico contro il sito di riconoscimento per MUC16 sono in grado di inibire fortemente l'adesione di cellule tumorali.

L'analisi della struttura secondaria e terziaria della mesotelina ha permesso di elaborare la struttura tridimensionale della proteina (Sathyanarayana *et al.*, 2009), risolvendo una configurazione a super-elica con motivi ripetuti di tipo-ARM, che a livello di struttura secondaria si ricollegano alle eliche in corrispondenza degli amminoacidi 313-317; 320-326; 329-334; 336-341; 346-358.

Precursore della mesotelina, otoancorina e stereocilina (tutte proteine ancorate alla membrana per mezzo di un'ancora GPI) possono essere ricondotte alla medesima superstruttura (Sathyanarayana *et al.*, 2009), supportando l'interpretazione secondo cui la mesotelina agirebbe da lectina a super-elica per il legame alla matrice glicoproteica extracellulare delle cellule cui è ancorata.

Il possibile coinvolgimento in processi di adesione cellulare è stato inoltre suggerito da studi di silenziamento genico per mezzo di un costrutto codificante per siRNA (short-interfering RNA). La trasfezione in linee cellulari di mesotelioma ha portato non solo ad una notevole riduzione di espressione della mesotelina, ma anche ad una diminuzione significativa nella capacità proliferativa delle cellule trattate (Wang *et al.*, 2012a).

La mesotelina ancorata alla membrana è una di due forme in cui è possibile ritrovare la proteina: nel siero è presente un secondo elemento (a peso molecolare di 42–45kDa) la cui sequenza aminoacidica all'N-terminale è uguale a quella della mesotelina di membrana. La forma solubile, che ha preso il nome di SMRP (soluble mesothelin related peptides), origina probabilmente a seguito di uno splicing alternativo non ca-

nonico tra gli esoni 16 e 17, che fa sì l'introne tra questi compreso venga inserito nel trascritto. L'evento comporta una mutazione per frameshift e l'introduzione di un codice di stop prematuro in corrispondenza dell'amminoacido 600, con formazione di una proteina trunca. Un possibile meccanismo alternativo di formazione della mesotelina solubile è il taglio proteolitico della mesotelina di membrana; in entrambe le ipotesi la proteina manca dell'estremità C-terminale coinvolta nell'ancoraggio alla membrana, e viene secreta (Hassan *et al.*, 2004).

Nei database genomici è infine possibile ritrovare una terza forma della mesotelina (isoforma 1, minoritaria) che si distingue per l'inserzione di 8 amminoacidi (PQAPRRPL) tra gli amminoacidi 408 e 409 dell'isoforma prevalente, ed a sua volta è ancorata alla membrana cellulare.

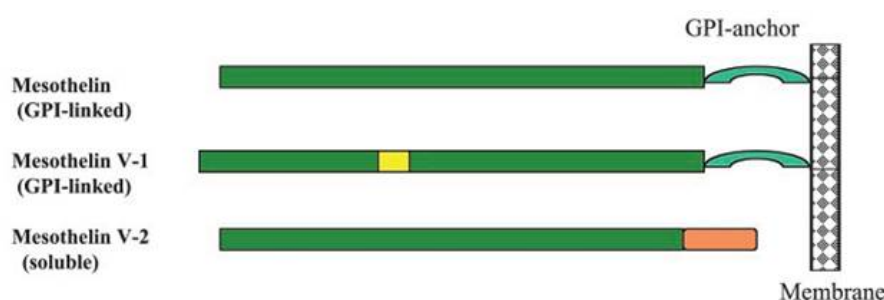


Figura 1.2: Rappresentazione schematica delle principali forme della mesotelina: MSLN (GPI-linked): isoforma 1, più comune; MSLN V-1: isoforma 2, con 8aa aggiuntivi; MSLN V-2: forma solubile SMRP.

A seconda del database genomico che si consulta, vengono riportate isoforme differenti per il messaggero della mesotelina, ma si possono riconoscere le tre forme già indicate Hassan *et al.* (2004). Secondo la classificazione operata da SwissProt:

- Isoforma 1; identifier Q13421-1. Questa isoforma, che corrisponde alla V-1 di Hassan Hassan *et al.* (2004) di lunghezza pari a 330 amminoacidi, è stata scelta dal browser come la sequenza “canonica” a cui fanno riferimento tutte le annotazioni, per quanto sia minoritaria.
- Isoforma 2; identifier Q13421-3. Identifica la mesotelina maggiormente rappresentata, a lunghezza di 322 amminoacidi poiché manca degli aa 409-416 della mesotelina V-1.
- Isoforma 3; identifier Q13421-2. Corrisponde alla mesotelina solubile SMRP, ed ha sequenza simile alla isoforma 2, dalla quale si differenzia per gli ultimi 30 amminoacidi.

- Isoforma 4; identifier Q13421-4. Rispetto alla isoforma 2, manca dell'amminoacido 44.

La scelta sull'isoforma da considerare come riferimento non si basa sull'incidenza (isoforma funzionale principale), bensì sulla lunghezza della variante, in modo che possa essere messo in evidenza un maggior numero di annotazioni.

Tutte le varianti riportate nel database genomico di Ensembl, UCSC, ed NCBI, sono illustrate in appendice A a pagina 71.

1.1.3 Promotore e regolazione della trascrizione

Il promotore del gene *MSLN*

Il promotore minimo del gene si colloca in una finestra di 317bps a monte del sito di inizio della trascrizione, ma mancano elementi caratteristici come le sequenze consensus per la TATA box e per la CAAT box (Urwin & Lake, 2000). Promotori TATA-less e CAAT-less sono solitamente tipici di geni housekeeping, e la loro assenza in un gene ad espressione tessuto-specifica è un dato poco comune e di non facile interpretazione. La trascrizione attivata del gene dipende da elementi a monte del promotore minimo, presumibilmente contenuti in una finestra di 1850bps e di 1285bps a monte del sito di inizio della trascrizione (Urwin & Lake, 2000).

Nonostante l'espressione strettamente tessuto-specifica, che possa realizzarsi anche in tessuti tumorali non mesoteliali suggerisce che la specificità non debba essere assoluta. A livello fisiologico, questa potrebbe essere raggiunta attraverso altri sistemi di regolazione indipendenti da elementi specifici del promotore, come la sua accessibilità nei diversi tessuti, o attraverso la presenza di un repressore tessuto-specifico non ancora identificato.

Nel 2007 è stata individuata una regione compresa tra i nucleotidi -164 e -78 che sembra in grado di attivare la trascrizione a partire dal promotore minimo in diverse linee cellulari tumorali che sovraesprimono la mesotelina, ma non in quelle che non la esprimono (Hucl *et al.*, 2007). Che poi l'espressione non si dimostri incrementata in linee cellulari derivanti da tessuti mesoteliali porta a pensare che in queste l'espressione sia tessuto-specifica piuttosto che cancro-specifica. All'interno di questa finestra sono emersi due elementi attivatori della trascrizione cancro-specifica, entrambi necessari per la funzione di enhancer: un sito tra i nucleotidi -82 e -75 ed uno tra i nucleotidi -94 e -85. Il primo sito (3' dell'enhancer) ha perfetta corrispondenza con l'elemento MCAT, mentre

il secondo sito, che si colloca al 5' dell'enhancer, ed è stato associato ad un sito di legame putativo per i fattori di trascrizione Sp1 e Sp1-like della famiglia Krüppel.

L'analisi comparata della sequenza dell'enhancer in diverse specie di mammifero (Huel *et al.*, 2007) ha rilevato un grado di conservazione che potrebbe supportare una effettiva funzionalità della regione identificata.

Oltre ad enhancer e siti di attivazione trascrizionale per il legame di fattori di trascrizione vi sono almeno altri due fattori che possono influire sui livelli trascrizionali di espressione di un gene: l'accessibilità del promotore e la presenza di mutazioni.

L'epigenetica

Per cambiamento epigenetico si intende un qualsiasi cambiamento ereditabile che non altera la sequenza nucleotidica di un gene, bensì la sua attività, ed infatti è stata definita da Arthur Riggs e colleghi come "lo studio dei cambiamenti mitotici e meiotici ereditabili che non possono essere spiegati tramite modifiche della sequenza di DNA" (Russo *et al.*, 1996).

Questi fattori non-genomici possono dunque agire in direzione di una diversa espressione dei geni dell'organismo, al punto che si ritiene oggi che sui fenomeni epigenetici si basi la maggior parte dei processi di differenziamento cellulare.

Tra i possibili meccanismi che possono provocare effetti epigenetici reversibili, ed in particolare il rimodellamento della conformazione della cromatina, si annoverano:

- modificazioni alle code N-terminali degli istoni;
- cambiamenti nei rapporti reciproci tra diversi nucleosomi;
- modificazioni del DNA, in forma di addizione covalente di gruppi metile a specifiche citosine (interne alle "isole-CpG") da parte delle metiltrasferasi.

Tutti questi processi, che difficilmente si presentano isolatamente ma mostrano una relazione reciproca (in particolare le modificazioni di istoni e DNA) alterano l'accessibilità fisica alle regioni del genoma sulle quali si legano proteine ed enzimi deputati all'espressione genica, alterandola.

La regione cromosomica in cui mappa il gene *MSLN*, circoscrivibile alle ultime 2Mb del braccio corto del cromosoma 16, è caratterizzata da un'alta percentuale di contenuto in GC, pari al 57,5% quando la media per il genoma umano è del 42% circa (Dunham *et al.*, 1999), e solo per il gene della mesotelina sono state predette 42 tra siti ed isole

CpG (Daniels *et al.*, 2001).

Tre regioni discrete nel promotore della mesotelina sono state caratterizzate come ipometilate in tessuti pleurici di soggetti affetti da mesotelioma pleurico maligno rispetto a pleure di pazienti sani, con una correlazione statisticamente significativa tra il livello di metilazione del promotore del gene ed i livelli di mesotelina sierica nei soggetti positivi per il mesotelioma (Nelson *et al.*, 2011).

Una analoga indicazione deriva da uno studio su 20 siti CpG, di cui 14 possibilmente rientranti in un'unica isola, e 6 a monte di questa (Tan *et al.*, 2010); la bassa percentuale di metilazione in 13 siti, per lo più appartenenti all'isola predetta, è risultata significativamente associata con alti livelli di espressione della mesotelina.

1.2 Genomica e genetica comparata

1.2.1 La genomica comparata

La genomica comparata rappresenta lo studio delle relazioni tra diverse specie per quel che ne concerne la struttura del genoma, l'organizzazione e le sue componenti; in astratto, può essere considerata come un tentativo di trarre vantaggio dalle informazioni registrate nel genoma quale risultato della pressione selettiva che su questi opera per comprendere l'azione ed i processi evolutivi che agiscono ed hanno agito sui genomi.

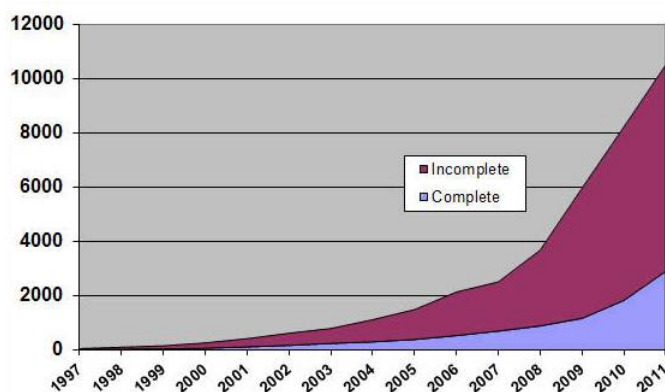


Figura 1.3: Il grafico illustra l'incremento nel tempo del numero di progetti di sequenziamento genomico, conclusi o in corso.

Negli ultimi anni, sotto la spinta del Progetto Genoma Umano, il numero di genomi a disposizione è cresciuto in modo quasi esponenziale (come mostra il grafico 1.3 nella pagina precedente), ed infatti ad oggi il database GOLD –Genomes Online Database– conta 6.869 progetti di sequenziamento genomico completati, 311 dei quali riguardano gli eucarioti; i progetti ancora in corso, o ultimati ma non pubblicati, sono invece 20.403, per un totale di 27.272 genomi indagati.

1.2.2 Applicazioni della genomica e della genetica comparata

Ora che la comparazione può essere effettuata tra interi genomi, molteplici sono le informazioni che possono essere reperite, a seconda del livello di distanza filogenetica scelto.

Alla base vi è sempre un allineamento di sequenze, che consiste nella mappatura dei nucleotidi di una sequenza in parallelo a quelli di una seconda (iterando il processo per allineamenti multipli) con l'introduzione di “buchi” (gap, “_”) laddove non vi è corrispondenza, al fine di incrementare il numero di posizioni con i medesimi nucleotidi tra le diverse sequenze.

Se comparando genomi di organismi molto distanti tra loro dal punto di vista filogenetico (divergenza ad 1 miliardo di anni fa o più), si possono ricavare indicazioni sulle tipologie funzionali dei geni presenti, in particolare su quelli che codificano per proteine e fattori alla base della funzionalità di una cellula o dell'organismo, a distanze più moderate (70–100 milioni di anni dalla divergenza) si possono riconoscere tanto sequenze funzionali quanto non funzionali all'interno del grande complesso di DNA conservato. In questo caso, le sequenze a valore funzionale mostrano i segni dell'azione di una selezione purificante, che fa sì queste mostrino un minor grado di divergenza rispetto al DNA non funzionale o a valenza neutrale (Jukes & Kimura, 1984). Infine, il confronto tra genomi molto simili, di specie che si sono separate solo di recente, è particolarmente indicato per individuare sequenze e regioni chiave per le differenze tra organismi, sottoposte a selezione positiva.

Quando si sceglie il livello filogenetico più appropriato per un determinato studio comparato bisogna tenere di conto del fatto che un livello di conservazione eccessivo può oscurare le regioni funzionali, ma al tempo stesso se vi è troppa divergenza queste ven-

gono nascoste. Così, un opportuno bilancio tra rilevanza biologica, distanza evolutiva ed analisi delle sequenze può fornire il supporto più solido per l'identificazione di sequenze conservate che sembrano evolvere sotto limitazioni evolutive in un background di sequenze che sono andate, per contro, incontro a divergenza per deriva genica (Nobrega & Pennacchio, 2004).

Per fare qualche esempio delle potenzialità dell'analisi genomica comparata, è possibile effettuare:

- Ricostruzioni filogenetiche ed evolutive; interi genomi di organismi tra loro relazionati vengono confrontati per ricostruire gli esatti rapporti di parentela tra specie.
- Inferenze sulle dinamiche genomiche; si basa sulla ricostruzione di eventi di duplicazione, inserzione, delezione, traslocazione e trasposizione di singoli geni, più geni o interi blocchi e frazioni di cromosomi.
- Predizione di elementi funzionali;
- Origine di particolari caratteristiche fenotipiche;

L'ultimo livello a cui si possono compiere analisi genetiche e genomiche comparate è quello di popolazione, tra individui diversi appartenenti alla medesima specie. Tre sono le principali tipologie di variazione che possono essere prese in considerazione:

- Variazioni strutturali nel genoma che coinvolgono blocchi di grandi dimensioni: grandi inversioni, inserzioni, delezioni, oppure fusioni);
- Copy number variations (CNVs), che comprendono duplicazioni e delezioni segmentali di interi cluster di geni o di singoli elementi;
- Restriction Fragment Length Polymorphisms (RFLP); micro inserzioni e micro delezioni, ristrette quindi a pochi nucleotidi, e polimorfismi a singolo nucleotide (SNPs).

Gli SNPs costituiscono la più grande fonte di variabilità tra individui, avendo una frequenza media di un polimorfismo ogni 600–1000 nucleotidi, e gli effetti di queste variazioni dipendono strettamente dal contesto in cui operano e dal tipo di sostituzione nucleotidica.

- Variable number of tandem repeats (VNTRs); si classificano come tali i minisatelliti ed i microsatelliti (STRs).

1.2.3 Omega

L'analisi genomica comparata si fonda su due parametri fondamentali: omologia, e differenze.

Entrambi sono alla base dell'analisi, per quanto tra gli obiettivi della stessa vi sia la loro individuazione, e questo si applica ad ogni livello di informazione, che sia la proteina, l'RNA, il DNA o le regioni regolatorie. Quello che si ricerca è la presenza di variazioni a carico del DNA, che vanno interpretate in termini di mutazioni sinonime e non-sinonime.

Le mutazioni sinonime sono quelle in cui cambia la sequenza nucleotidica, ma a livello funzionale le proprietà della proteina non variano, poiché il codone mutato codifica sempre per il medesimo amminoacido, mentre nelle mutazioni non-sinonime si ha il cambiamento di un amminoacido, con conseguente alterazione nella struttura e potenzialmente nella funzionalità della proteina.

Le mutazioni sinonime potrebbero riflettere il ritmo di base con cui il genoma muta, ed hanno tasso di introduzione per lo più costante: ogni volta che avviene tale mutazione, non ne deriva né una conseguenza positiva né una negativa, per cui possono fissarsi o perdersi senza che vi intervenga un meccanismo selettivo in favore o contro.

Queste due categorie di variazioni (sinonime e non) sono messe in relazione l'una con l'altra, così da ottenere il valore ω (omega), dato dal rapporto tra i due tassi di sostituzione.

Una volta ottenuto l'allineamento, si procede con la conta delle mutazioni puntiformi sinonime (N_s) e non sinonime (N_a); dal momento che le sequenze delle diverse specie potrebbero non essere di identica lunghezza, ogni valore ottenuto deve essere diviso per la lunghezza totale della relativa sequenza.

I risultati ottenuti dal rapporto tra il numero di sostituzioni ed il numero totale dei siti per ciascuna specie vengono indicati come K_a (per le non sinonime) e K_s (per le sinonime). K_s riflette il tasso evolutivo di base, riferendosi alle mutazioni sinonime, supposte neutrali, mentre K_a è relativo al tasso evolutivo selettivo; è il rapporto (K_a/K_s) a fornire l'indice ω , che riflette il tasso evolutivo (relativo) della sequenza:

- $K_a \approx K_s$; bassa se non nessuna selezione ha agito tra siti differenti
- $K_a < K_s$; una forte selezione purificante ha ridotto il tasso di fissazione di mutazioni deleterie

- $Ka > Ks$; il tasso di fissazione di cambiamenti aminoacidici è maggiore rispetto al tasso di sostituzione neutrale, ed è quindi segno di una selezione positiva per sostituzioni aminoacidiche che conferiscono un vantaggio selettivo all'organismo (Emes, 2003).

Nei genomi Ks è molto maggiore rispetto a Ka quando il raffronto è fatto a livello di popolazione, ma la situazione può cambiare sensibilmente quando si fa la proporzione tra specie differenti, e questo perché sono le mutazioni non sinonime le maggiori candidate quali responsabili dei processi evolutivi tra le specie.

Per finire, anche in questo caso si può scegliere il livello a cui calcolare ω , o su tutto il genoma o su singole sequenze a seconda dello scopo dell'analisi. Quando ω viene calcolato sull'intero genoma, rappresenta un buon indicatore delle strategie evolutive che le diverse specie o gruppi di specie hanno sviluppato nei confronti della selezione e del rapporto con l'ambiente sperimentato (di norma, più alto è Ka e più rilassata è la selezione negativa).

Se calcolato su un gene o su una regione più o meno circoscritta, ω suggerisce se di recente il gene ha subito un'accelerazione evolutiva; in questo caso il rapporto Ka/Ks viene calcolato come la media tra tutti i codoni in un gene. Tuttavia, l'adattamento molecolare di una proteina sottoposta a selezione positiva si realizza solo a livello di pochi, specifici residui e per lo più solo in certe linee evolutive (Emes, 2003). Si consideri un enzima, paralogo od ortologo ad un altro: i residui coinvolti nel riconoscimento e selezione del substrato saranno soggetti a selezione positiva (alta incidenza di mutazioni non-sinonime), mentre tutti gli altri domini, in particolare il sito catalitico, saranno sottoposti a selezione purificante in merito al vincolo funzionale. Per questo, il riscontro di selezione positiva può essere conseguito unicamente calcolando il rapporto Ka/Ks a livello sito-specifico, e non sull'intera sequenza.

Un elemento che non si può trascurare quando si studiano singoli geni è l'incomplete lineage sorting (ILS), un fenomeno di assortimento incompleto dei loci per cui, tra diverse linee evolutive aventi un antenato comune, i polimorfismi presenti nella popolazione ancestrale comune a tutte possono essersi perse in una specie e mantenute in altre, indipendentemente dalle dinamiche di divergenza della specie. Come conseguenza, la ricostruzione filogenetica basata su alcuni geni o loci potrebbe portare ad una rappresentazione diversa da quella che altrimenti si otterrebbe dall'analisi comparata sull'intero

genoma (Hobolth *et al.*, 2011).

1.2.4 La mesotelina umana ed i suoi ortologi

Nella sezione 1.1.2 a pagina 3 (Struttura della Mesotelina) sono state brevemente introdotte le principali caratteristiche del gene *MSLN* umano, che si compone di 17 esoni, per un totale di 8101 nucleotidi (NC_000016.9).

Per l'analisi filogenetica del gene *MSLN* sono state scelte 17 specie di mammifero (16 placentati ed un marsupiale che è stato imposto come outgroup) sulla base di criteri descritti in appendice C.1 a pagina 101; i principali dettagli del gene per ciascuna specie sono riportati in tabella 1.1 a fronte.

L'analisi qui condotta si prefigge di valutare la mesotelina a livello di sequenza codificante ed aminoacidica, così da individuare se vi sono particolari siti che in una o più linee evolutive hanno accumulato variazioni in grado di determinare differenze strutturali o funzionali della proteina, magari a carico di regioni coinvolte nell'interazione proteica. Lo studio è subordinato e complementa una indagine complessivamente volta ad individuare siti importanti per la regolazione post-trascrizionale di espressione della mesotelina, che sembra essere piuttosto complessa anche in virtù della sua valenza in contesto fisiologico e tumorale.

A tal fine, per l'RNA messaggero è stato testato il rapporto Ka/Ks a due livelli di complessità, valutando le sequenze per intero e poi ciascun esone singolarmente, così che gli eventuali segnali evolutivi non venissero mascherati.

Tabella 1.1: Caratteristiche del gene nelle specie scelte per l'analisi.

Specie	Database	#Esoni	DNA		cDNA		Proteina	
			Accession	Dimensioni	Accession	dimensioni	Accession	dimensioni
<i>Homo sapiens</i> (Human)	NCBI	17	NC_000016.9	8101	NM_005823.5	2187	NP_005814.2	622
<i>Pan paniscus</i> (Bonobo)	NCBI	17	NW_003870224.1	8043	XM_003808819.1	2162	XP_003808867.1	622
<i>Nomascus leucogenys</i> (Gibbon)	NCBI	18	NC_019833.1	7830	XM_004091295.1	2160	XP_004091343.1	622
<i>Macaca mulatta</i> (Rhesus)	NCBI	16	NC_007877.1	5668	XM_001087333.2	1988	XP_001087333.2	621
<i>Papio anubis</i> (Baboon)	NCBI	18	NC_018171.1	7568	XM_003916301.1	2158	XP_003916350.1	622
<i>Saimiri boliviensis</i> (Squirrel Monkey)	NCBI	18	NW_003943633.1	9621	XM_003928359.1	2086	XP_003928408.1	623
<i>Mus musculus</i> (Mouse)	NCBI	16	AC_000039.1	5715	NM_018857.1	2135	NP_061345.1	625
<i>Rattus norvegicus</i> (Rat)	NCBI	16	AC_000078.1	5662	NM_031658.1	2099	NP_113846.1	625
<i>Cricetus griseus</i> (Hamster)	NCBI	16	NW_003613852.1	5566	XM_003501524.1	1917	XP_003501572.1	626
<i>Sus scrofa</i> (Pig)	Ensembl	10	ENSSSCG00000008001.2	3354	ENSSSCT000000008769	1497	ENSSSCP000000008545	403
<i>Bos taurus</i> (Cow)	NCBI	10	NC_007326.5	3584	NM_001100374.1	1458	NP_001093844.1	403
<i>Equus caballus</i> (Horse)	Ensembl	11	ENSECAG00000001261.1	3221	ENSECAG00000001261	1206	ENSECAP00000000931	401
<i>Felis catus</i> (Cat)	NCBI	10	NC_018738.1	3192	XM_003999055.1	1206	XP_003999104.1	401
<i>Mustela putorius furo</i> (Ferret)	Ensembl	11	ENSMPUG00000015901.1	4133	ENSMPUT000000016036	2305	ENSMPUP000000015795	401
<i>Canis lupus familiaris</i> (Dog)	NCBI	10	NC_006588.3	3189	XM_849019.2	1300	XP_854112.1	401
<i>Odobenus rosmarus divergens</i> (Walrus)	NCBI	10	NW_004450442.1	3071	XM_004403103.1	1300	XP_004403160.1	399
<i>Monodelphis domestica</i> (Opossum)	Ensembl	28	ENSMODG00000016813	39084	ENSMODT000000021355	3953	ENSMODP000000020985	1202

1.3 I microRNA

I microRNA (miRNA) sono piccole molecole di RNA non codificante endogene a singolo filamento, lunghi 22nt circa, che agiscono come regolatori dell'espressione genica a livello post-trascrizionale in metazoi, protozoi e piante (Fabian *et al.*, 2010).

1.3.1 Biogenesi dei microRNA

I geni codificanti per microRNA possono trovarsi o come unità indipendenti o all'interno di altri geni, tanto negli esoni quanto negli introni (Rodriguez & Griffiths-Jones, 2004) o in entrambi a seconda dello splicing alternativo dei trascritti, orientati con la medesima direzione del trascritto o sul filamento antisense (Bartel *et al.*, 2004). Oltre a singoli geni codificanti per microRNA, nel genoma è possibile ritrovare interi cluster il cui pattern di espressione ed organizzazione suggerisce un sistema di trascrizione multi-cistronico (Lagos-Quintana *et al.*, 2001). All'interno di un cluster, i microRNA sono generalmente coinvolti nella regolazione del medesimo processo biologico (dove interagiscono tuttavia con messaggeri differenti), con una relazione funzionale piuttosto che di sequenza.

I microRNA che costituiscono unità trascrizionali autonome, o che sono trascritti sul filamento antisense di altri geni, presentano un promotore e/o enhancer che ne regola la trascrizione, svolta dalla RNA polimerasi-II (Lee *et al.*, 2004). Per contro, microRNA che risiedono all'interno degli introni di altri geni codificanti vengono co-trascritti con l'ospite, e la loro escissione dal pre-mRNA è operata da una esonucleasi (Tycowski *et al.*, 1993).

La maggior parte dei microRNA mostra un profilo di espressione spatio-temporale molto preciso, al punto da poter riconoscere come nei metazoi molti tipi cellulari, in ciascun stadio di sviluppo, presentino un distinto profilo di espressione dei microRNA, che può essere sensibilmente alterato in contesto patologico (Bartel *et al.*, 2004).

Alla trascrizione fanno seguito diversi step di maturazione del microRNA:

Maturazione nucleare Il trascritto primario (pri-miRNA) può avere lunghezza molto eterogenea, ma si basa sempre su una struttura a forcina (o più) di circa 70nt fiancheg-

giata da sequenze necessarie per l'efficiente processamento del precursore. A livello nucleare, la struttura a doppio filamento della forcina viene riconosciuta dalla proteina DGCR8, che si associa alla RNasi-III Drosha e ad altri cofattori (Han *et al.*, 2004). Drosha effettua un taglio specifico a circa 11nt dal segmento basale della forcina, lasciando una struttura a stem e loop con 5'-fosfato e 3' overhang che prende il nome di pre-miRNA (Basyuk *et al.*, 2003).

Il pre-miRNA viene quindi traslocato nel citoplasma grazie all'esportina 5, che in associazione a RAN-GTP supera il complesso del poro nucleare (Yi *et al.*, 2003; Lund *et al.*, 2004).

Maturazione citoplasmatica Una volta nel citosol, il pre-miRNA viene riconosciuto da una seconda RNasi-III, Dicer, che riconosce le estremità 5'P e 3'overhang lasciate da Drosha ed effettua un taglio alla base del loop terminale, dopo due giri di elica dalla base dello stem. Dicer è associata al partner proteico TRBP (Trans-activator RNA binding protein) e potenzialmente ad una proteina della famiglia Argonaute (Diederichs & Haber, 2007). Quello che si ottiene è un double-strand microRNA di circa 22nt di lunghezza che presenta ad entrambi gli estremi del duplex estremità protrudenti 5'P e 3'overhang, indicato come miRNA:miRNA*.

Solo uno dei due filamenti del duplex, definito "guida" (il miRNA), viene caricato sul complesso ribonucleoproteico RISC (RNA-Induced Silencing Complex) (Salzman *et al.*, 2007), mentre il filamento "passeggero" (miRNA*) viene rilasciato dal complesso e generalmente degradato (Schwarz *et al.*, 2003), anche se vi sono evidenze di casi in cui entrambi i filamenti del duplex sono parimenti ritenuti e caricati sul complesso RISC per la regolazione di sottogruppi di mRNA differenti. In questa circostanza le due specie risultanti vengono identificate come -5p o -3p a seconda del braccio della forcina da cui derivano (Chiang *et al.*, 2010).

Si ritiene che la specie prevalente, quella che viene scelta preferenzialmente per entrare nel complesso RISC, è quella che nel duplex ha al 5' energia di legame minore (Schwarz *et al.*, 2003).

La specificità del primo taglio operato da Drosha sembrerebbe quindi responsabile del registro di taglio all'interno del precursore, e quindi della definizione di entrambe le estremità del duplex maturo (Lee *et al.*, 2003).

Tra i principali componenti del complesso RISC vi sono le proteine Argonaute (Ago),

che interagiscono direttamente con il microRNA, le proteine GW182, che interagiscono con Ago ed agiscono a valle di questa, effettuandone la funzionalità regolatoria, ed altre proteine, che possono agire come fattori di regolazione modulanti l'azione dei microRNA.

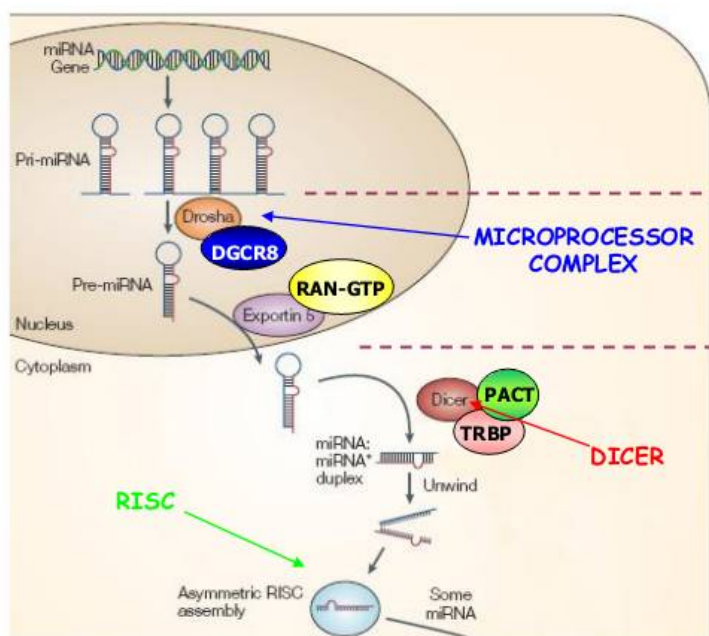


Figura 1.4: Raffigurazione schematica del processo di maturazione dei microRNA, dalla trascrizione del gene fino all'assemblaggio del complesso RISC

Recentemente sono stati scoperti microRNA che presentano un pathway di biogenesi non-canonica: i “mirtroni” ed i “simtroni”; i geni codificanti ricadono in entrambi i casi all'interno di introni di geni ospiti, ma se per i primi il processamento è dipendente dallo splicing (Havens *et al.*, 2012), e coinvolge un Lariat debranching enzyme al posto di Droscha, nei secondi la maturazione nucleare richiede Droscha ma non DGCR8, così come non sembrano necessari l'esportina 5 o Dicer e le Argonaute a livello citoplasmatico.

1.3.2 Principi di interazione microRNA-mRNA

I microRNA hanno la funzione di veicolare il complesso RISC all'RNA messaggero, determinando la specificità di legame. Il riconoscimento si fonda sull'appaiamento

tra basi complementari tra il microRNA e l'mRNA; per lo più, nei metazoi l'appaiamento è imperfetto, mentre nelle piante si ha un appaiamento complementare per tutta la lunghezza o quasi del miRNA (Rhoades *et al.*, 2002).

Nei metazoi, uno dei requisiti più stringenti per la regolazione è il perfetto e contiguo appaiamento dei nucleotidi 2–8 al 5' del miRNA, che rappresenta la seed region di nucleazione con l'mRNA.

Quando più siti di interazione con microRNA sono presenti sullo stesso messaggero nella 3'UTR, si osserva una maggior destabilizzazione del trascritto, in accordo ad un effetto moltiplicativo risultato di azione indipendente e cooperativa (Grimson *et al.*, 2007); nel caso di due match site in stretta prossimità (8–40nt circa), la repressione risulta però maggiore di quanto non si avrebbe se i due siti contribuissero in modo indipendente.

L'analisi comparata tra siti di nucleazione in UTR ortologhe di mammifero ha rivelato un significativo arricchimento in match-site doppi a poca distanza l'uno dall'altro, che potrebbe sottendere ad una preferenzialità biologica per questa situazione.

Oltre all'appaiamento del seed, sembra che anche un appaiamento perfetto tra i nucleotidi 13–16 possa avere un ruolo nell'efficacia della regolazione (Grimson *et al.*, 2007), ed infatti questi sono i nucleotidi più conservati al di fuori del seed.

Più microRNA possono legarsi a, e quindi regolare, lo stesso messaggero, ed allo stesso tempo più messaggeri diversi possono essere il target del medesimo microRNA.

1.3.3 Modalità di azione

Quando fu scoperto il primo microRNA, lin-4 di *C.elegans*, ed il suo meccanismo di azione fu analizzato, emerse una significativa riduzione della proteina-bersaglio lin-14, mentre i livelli del suo messaggero venivano poco influenzati dalla presenza o assenza del microRNA (Wightman *et al.*, 1993). I dati furono interpretati come evidenze di una repressione a livello post-trascrizionale successiva all'inizio della traduzione; in alternativa, fu anche proposto che piuttosto che sulla traduzione, la repressione sarebbe stata esplicata sul polipeptide nascente, in forma di degradazione proteolitica selettiva (Olsen & Ambros, 1999).

Anni dopo, Petersen *et al.* (2006) hanno proposto un modello secondo cui i microRNA agirebbero da antagonisti della fase di allungamento durante la sintesi proteica, determi-

nando una prematura terminazione e distacco dei ribosomi; analogamente, altri autori suggerirono che il complesso miRISC potrebbe richiamare enzimi proteolitici in prossimità di ribosomi in attiva traduzione (Nottrott *et al.*, 2006) (le due possibili modalità sono illustrate in fig. 1.5).

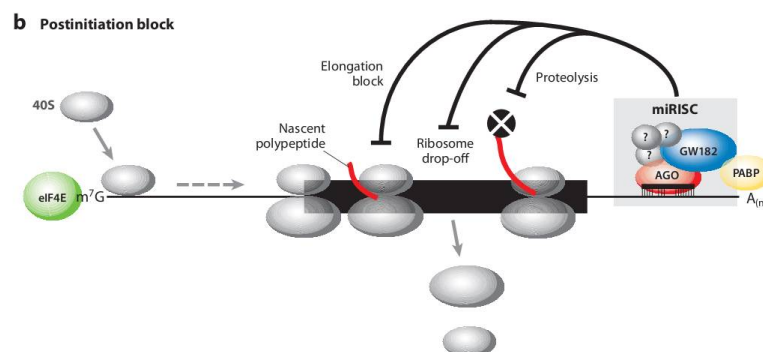


Figura 1.5: Raffigurazione schematica delle due possibili modalità di repressione post-trascrizionale dei microRNA, successiva alla fase di inizio della traduzione: blocco dell'allungamento e degradazione proteolitica del polipeptide nascente.

Nel giro di pochi anni cominciarono ad emergere indicazioni di processi di regolazione differenti: nelle piante, Rhoades *et al.* (2002) misero in luce un processo noto come “slicing”, in cui l'assenza di proteina si spiega in virtù della degradazione del messaggero da parte del complesso RISC. La differenza fu imputata al grado di complementarietà tra repressore e bersaglio: ad un appaiamento quasi perfetto corrisponderebbe la capacità del complesso RISC nell'operare un taglio nel messaggero (primo passo alla completa degradazione), mentre ad un appaiamento limitato a pochi nucleotidi del microRNA una inibizione di tipo traduzionale (Bartel *et al.*, 2004).

Nuove evidenze a supporto di modalità di inibizione della traduzione differenti emersero anche per i metazoi:

Repressione della fase di inizio della traduzione

Il complesso miRISC sembrerebbe in grado di inibire l'avvio della traduzione interferendo con il riconoscimento del cap al 5' da parte di eIF4F, ostacolando così il reclutamento della subunità 40S del ribosoma al messaggero. In alternativa, il complesso

potrebbe interferire con l'unione della subunità 60S, ma in ogni caso verrebbe prevenuta la formazione del complesso ribosomiale 80S (Fabian *et al.*, 2010).

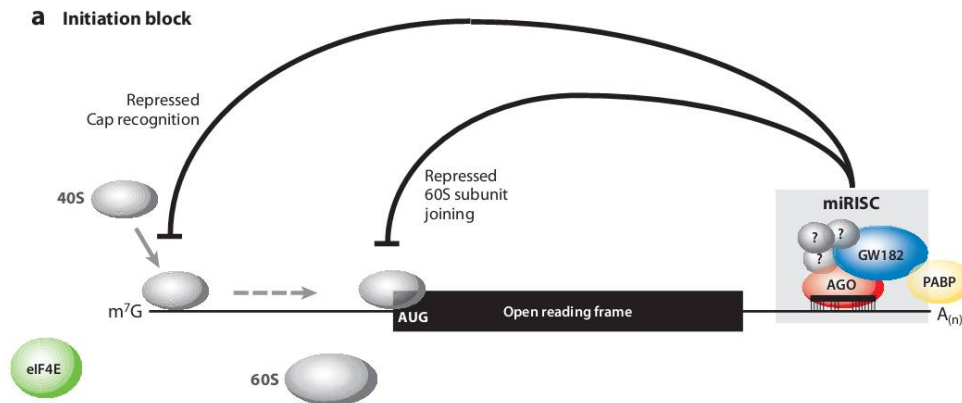


Figura 1.6: Rappresentazione schematica delle due possibili modalità di repressione post-trascrizionale dei microRNA, precedente alla fase di inizio della traduzione: ostacolo del riconoscimento del cap al 5' da parte della subunità 40S ed ostacolo al legame della subunità 60S.

È stato inoltre riportato (Fabian *et al.*, 2009) che l'interazione di GW182 con le PABPs (PoliA binding protein) potrebbe interferire con la conformazione chiusa circolarizzata (mediata dall'interazione tra eIF4G e PABP) che assumono i messaggeri al momento della traduzione, contribuendo ulteriormente alla repressione dell'inizio della traduzione.

Promozione della traduzione o della trascrizione

Il miR-10a è capace di interagire con la 5'UTR di alcuni mRNA che codificano per proteine ribosomali, determinandone un aumento del rate traduzionale in condizioni di stress o di carenza nutrizionale (Orom *et al.*, 2008).

Oltre alla 5'UTR, vi sono evidenze di microRNA che legano il promotore di geni e ne incrementano la trascrizione, come nel caso del miR-373, che nell'uomo può associarsi al promotore dei geni *E-caderina* e *CSDC2*, in un contesto cellulare-specifico (Place *et al.*, 2008).

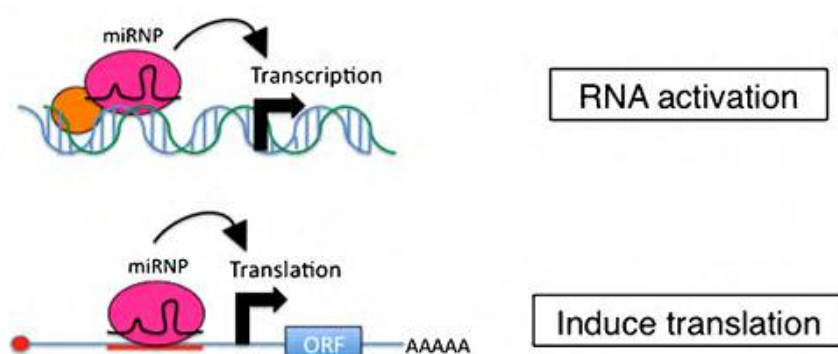


Figura 1.7: Raffigurazione schematica delle due possibili modalità di attivazione trascrizionale dei microRNA: legame al promotore del gene a livello nucleare ed interazione con la 5'UTR del messaggero.

L'aspetto più critico di questo modello riguarda il suo contesto di azione, che da citoplasmatico si sposta a nucleare: al momento, solo AGO2 è stato identificato come possibile responsabile della localizzazione nucleare di particolari RNA (Guang *et al.*, 2008), per quanto non sia ancora stato delucidato come questa proteina e gli RNA legati vengano importati nel nucleo.

Degradazione del messaggero

Dal 2005, diversi studi hanno messo in evidenza come in diversi gruppi filetici (uomo incluso) i microRNA possono indurre una significativa degradazione del messaggero target anche in condizioni di complementarietà imperfetta (Bagga *et al.*, 2005; Wu & Belasco, 2005; Mishima *et al.*, 2006).

A supporto, furono mostrati risultati di esperimenti in cui la sovra-espressione di microRNA in cellule in cui non sono normalmente espressi porta ad una riduzione dei trascritti che presentano siti di legame per questi (Lim *et al.*, 2005; Selbach *et al.*, 2008).

Nelle cellule animali, le proteine Argonaute ed il complesso miRISC in generale non sono in grado di degradare l'mRNA per mezzo di tagli endonucleolitici –come succede invece per lo slicing nelle piante, salvo nel caso in cui il miRNA è del tutto complementare al target– ma piuttosto dirigono il messaggero al macchinario generale di degradazione. L'interpretazione ha trovato il supporto di studi nei quali i microRNA sono risultati in grado di promuovere i processi di deadenilazione e decapping dei corrispettivi target (Behm-Ansmant *et al.*, 2006; Mishima *et al.*, 2006; Eulalio *et al.*, 2007).

Il processo di deadenilazione sembra un effetto generalizzato della regolazione ad opera

dei microRNA (al punto da essere indipendente dalla trascrizione del messaggero), e per quanto non rappresenti una caratteristica necessaria ed indispensabile al silenziamento, quando avviene, l'inibizione risulta incrementata (Eulalio *et al.*, 2009).

Oltreché reclutare il macchinario di decadimento dell'mRNA, GW182 interagisce con fattori di trascrizione e/o subunità ribosomali presenti a livello del messaggero per ostacolarne la trascrizione; tra gli interattori evolutivamente conservati di GW182 si sono già citate le PABPs, e l'interazione tra le due componenti sembrerebbe avere come conseguenza il portare la coda di poliA –a cui sono associate le PABPs– in prossimità del complesso di deadenilasi associato a miRISC, facilitando la reazione di rimozione della coda di poliA (Fabian *et al.*, 2009).

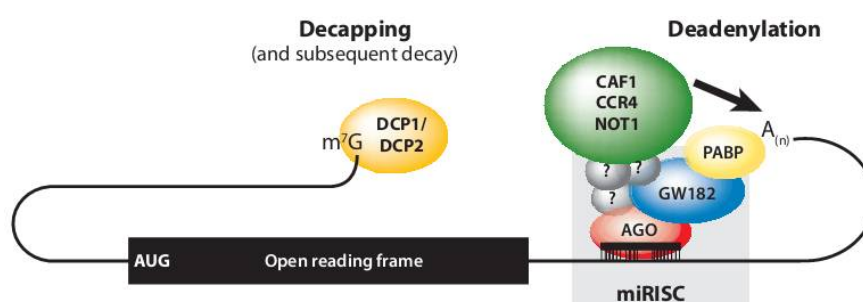


Figura 1.8: Modalità di destabilizzazione del messaggero che potrebbero stare alla base del processo di degradazione da parte del complesso miRISC: rimozione del cap al 5' e/o rimozione della coda poliA.

L'interpretazione al momento più supportata è quella che il messaggero sia prima soggetto a deadenilazione, quindi a decapping ed infine a degradazione ad opera di esonucleasi che tagliano dal 3' al 5'. È stato anche avanzato un modello alternativo, in cui i messaggeri subiscono il de-capping al 5' da enzimi specifici del complesso DCP1-DCP2, e secondariamente la degradazione 5'–3' da parte della esonucleasi Xrn1, senza deadenilazione (Coller & Parker, 2004).

Sono state riportate evidenze di microRNA che possono determinare la degradazione di messaggeri senza che fossero in attiva traduzione (Mishima *et al.*, 2006; Wu *et al.*, 2006), ad indicare che la degradazione è indipendente dallo stato di traduzione, e suggerendo che questa non sia la sola conseguenza della repressione traduzionale esplicita dai microRNA.

Legame alla regione codificante

La seconda eccezione al “dogma” secondo cui i microRNA si legano con il complesso miRISC alla 3’UTR di messaggeri target per reprimerne l’espressione viene dalla scoperta di microRNA capaci di legarsi con appaiamento imperfetto alla regione codificante del messaggero, reprimendone anche in questo contesto l’espressione (Tay *et al.*, 2008). Questi microRNA sembrano però incrementare l’effetto regolatorio di microRNA che interagiscono con la 3’UTR secondo un effetto sinergico (Fang & Rajewsky, 2011), piuttostochè esplicare una repressione autonoma.

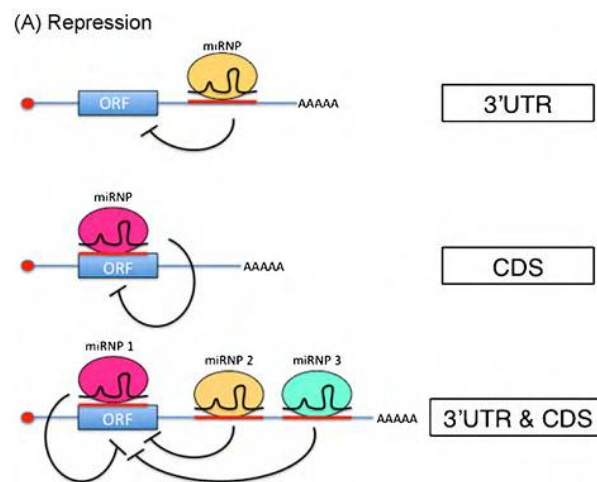


Figura 1.9: Principali modalità di legame e di regolazione della traduzione da parte dei microRNA: legame alla 3’UTR, legame alla ORF e legame di più microRNA a diversi siti target

Sembra quindi che si sia sviluppato un network di microRNA che agiscono simultaneamente così da conseguire una regolazione più fine dell’espressione genica e proteica, legandosi a più regioni di complementarietà per microRNA differenti tanto all’interno della regione codificante quanto della 3’UTR sullo stesso messaggero target.

1.3.4 Considerazioni evolutive

L’azione regolatoria dei microRNA è stata ampiamente documentata in entrambi i regni di piante e metazoi, e due sono le ipotesi, contrastanti, riguardo la loro origine: accettando l’interpretazione secondo cui piante e metazoi discendono da un ultimo an-

tenato comune unicellulare, si deve pensare ad una convergenza evolutiva piuttosto che ad una eredità (Chen & Rajewsky, 2007), per quanto l'evoluzione sia stata precoce in entrambe le linee. Vi sono tuttavia evidenze di elementi simili a quelli che fanno parte oggi del macchinario proteico operante nel processo di RNA-interferenza già presenti nell'ultimo ancestore comune a tutti gli eucarioti Shabalina 2008. Di fatto, si suppone che un meccanismo di difesa dai parassiti assimilabile a quello dei siRNA (short interfering RNA) fosse stato sviluppato già dai primissimi eucarioti, e non si può quindi escludere l'analoga presenza di un sistema di controllo mediato da "proto-microRNA".

Quando si guarda a famiglie geniche per microRNA omologhe in linee evolutive distanti, si può osservare come queste siano andate incontro ad evoluzione in parallelo, alcune per contrazione ed altre per espansione dei paraloghi (Bartel *et al.*, 2004), essendo caratterizzate da processi evolutivi identici a quelli di geni codificanti per proteine (duplicazione in tandem, duplicazione segmentale o dell'intero genoma), interpretazione sostenuta dal fatto che il 37% dei geni codificanti per microRNA siano organizzati in cluster (Altuvia *et al.*, 2005).

Si suppone che, alla nascita di un nuovo microRNA, questo sia capace di legarsi in modo non specifico e casuale ad una molteplicità di messaggeri, dal momento che il seed minimo è solo di pochi nucleotidi. Per evitare effetti deleterei, il loro livello di trascrizione e di espressione iniziale sarebbe quindi molto basso (Chen & Rajewsky, 2007) e solo con il tempo, se il ruolo del nuovo microRNA si rivela fisiologicamente significativo, l'effetto della selezione purificante farà sì che vengano rimossi eventuali siti di off-target nel genoma, mantenendo al contempo quelli utili. Solo in ultimo si assiste all'incremento di espressione del microRNA.

1.3.5 Il microRNA miR204-5p

Mediante la consultazione di tools bio-informatici è possibile ritrovare l'elenco di molteplici microRNA che hanno la potenzialità di legare la mesotelina, tuttavia nessuno al momento è stato validato da studi funzionali, né in vivo né in vitro, con le predizioni che rimangono quindi a livello di affidabilità computazionale.

Per la scoperta di microRNA potenzialmente coinvolti in pathway specifici, l'approccio al momento più sfruttato è quello di analisi con piattaforme di microarray per ottene-

re un quadro completo dei livelli di espressione di molteplici microRNA presenti nello specifico tessuto o momento di sviluppo, a seconda delle finalità. Una applicazione che si è rivelata molto informativa, almeno a livelli preliminari di indagine, è quella del confronto tra profili di espressione dei microRNA in contesti patologici e fisiologici, dal momento che l'individuazione di microRNA differenzialmente espressi può aiutare ad individuare elementi importanti per la determinazione dei corrispettivi fenotipi. Questo tipo di studi ha di fatto contribuito a confermare che diversi tumori mostrano ciascuno un profilo specifico di espressione dei microRNA, al punto che alcuni possono assumere valore diagnostico e/o prognostico (Gottardo *et al.*, 2007; Enerly *et al.*, 2011; Guled *et al.*, 2009).

Il microRNA miR-204-5p è stato selezionato a seguito del confronto di lavori di microRNA profiling high throughput in tessuti specifici (pancreas, ovario e pleura, in cui la mesotelina è sovraespressa nel fenotipo tumorale) tra soggetti sani ed affetti da tumore, al fine di individuare quelli differenzialmente espressi. Tra i microRNA identificati, sono stati selezionati quelli in comune a tutti i lavori, e tra questi il miR-204-5p si è rivelato il solo predetto come capace di legare la mesotelina a livello di regione codificante, ed è quindi stato scelto per lo studio funzionale in vitro (per tutti i dettagli, fare riferimento alla sezione apposita nell'appendice B.1 a pagina 75).

Nell'uomo, il miRNA-204-5p (MIMAT0000265 del database miRBase) è uno dei prodotti di maturazione del gene miR-204, che mappa sul cromosoma 9 (73424891–73425000), e viene trascritto sul filamento antisenso dal processamento intronico del gene TRPM3 (Transient receptor potential cation channel, subfamily M, member 3). Tra i target validati di questo microRNA vi sono più di 30 messaggeri (TARBASE), ed ancor più sono quelli predetti, ma non la mesotelina.

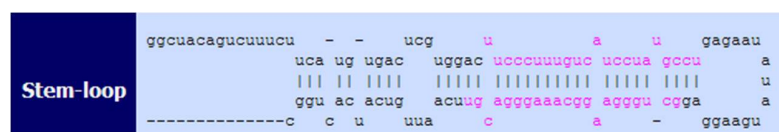


Figura 1.10: Struttura a forcina del precursore codificante per il miR-204-5p

Il miR-204-5p maturo ha sequenza: UUCCCUUUGUCAUCCUAUGCCU, ed è perfettamente conservato in tutti i vertebrati di cui si dispone attualmente di un genoma sequenziato, parzialmente o per intero (<http://blast.ncbi.nlm.nih.gov>). Si può di conseguenza inferire

che la sua funzione regolatoria sia ben consolidata sotto il profilo evolutivo, come si può apprezzare dalla fig. 1.11 nella pagina successiva.

I programmi per la predizione di possibili siti di legame di un microRNA ad un trascritto sono capaci di fornire indicazioni piuttosto accurate, ma si basano comunque su algoritmi che interpretano l'output in termini di energie di legame tra sequenze (ΔG).

Dal momento che nel mio progetto di tesi si valuta anche l'aspetto filogenetico ed evolutivo delle sequenze, ho voluto indagare il grado di conservazione nel contesto del microRNA miR204-5p: l'obiettivo è valutare il livello di omologia dei siti predetti di nucleazione tra microRNA e messaggero target in 17 specie di mammifero. L'approccio prevede di eseguire per ciascuna una analisi indipendente come fatto a livello preliminare per *Homo sapiens*, ed unire in un unico file di allineamento multiplo i risultati, per vedere se vi è almeno un seed match per il miR-204-5p conservato in tutte le specie. In tal caso, questo sarebbe un ulteriore e valido supporto non solo all'effettiva potenzialità di interazione, ma potrebbe fornire anche un'indicazione sul sito da trattare come di elezione per il targetting.

```

Homo      --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Pan       --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Gorilla   --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Pongo     --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Nomascus  -----GCCAAT---GATGACAAATTGAACG TCCCTTTGCTCTCCAGCCTCTTCATATATTCT--CAGGCATAGGATGACAAAGGGACGT
Callithrix --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Macaca    --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Equus     --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Otolemur  -----GCCAGT---GATGACAAATTGAACG TCCCTTTGCTCTCCAGCCTCTTCATATATTCT--CAGGCATAGGATGACAAAGGGACGT
Felis     --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Canis     -----TGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Oryctolagus -----TGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Sus       --GGCTA-CAGTCTTCT-CATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Mus       -----TGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Rattus    --GGCTACAGCCCTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Bos       --T-----TTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Macropus  --GACGACAGTCTCTCCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Monodelphis --GACT---ACAGTCTCCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Pelodiscus --TGCTACAATATCTATTCTCATGTGACCCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Sarcophilus --GGCTACCGTCTCTCCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Dipodomys -----GCCAGT---GAT---CAATTGAACG TCCCTTTGCTCTCCAGCCTCTTCATATATTCT--CAGGCATAGGATGACAAAGGGACGT
Anolis    --GTCTGCACTACCTATTCATGTGACCCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Ornithorhynchus --GACTACTATCTCTCTCCATGTGACCCGTGGGCT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Taeniopygia --GTCAGCAGTATCTGTTTCATGTGACCCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Xenopus   -----TGTGACCTGTGGGCT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Latimeria --T-----CTGTGCATGTGGCCCGTGGGCT TCCCTTTGTCATCCTATGCCTGAGATTAAATACAAA--AGGGGCTGGGAAGGCAAGGGACGT
Oreochromis -----TGTGACCTATGGACT TCCCTTTGTCATCCTATGCCTGGAATCAAATAAAGGGGCTGGGAAGGCAAGGGATGC
Oryzias   --T-----ATGACATGTGGACT TCCCTTTGTCATCCTATGCCTGGA--GTAATAGA--GGGGCTGGGAAGTCAAAGGGACGC
Danio     --T-----CATGTGACCTGTGGACT TCCCTTTGTCATCCTATGCCTGGA--GTAATAGA--GGGGCTGGGAAGTCAAAGGGACGC
Xiphophorus --T-----GTGGCCTGTGGACT TCCCTTTGTCATCCTATGCCTGGAGTCAATGTCAA--TGGGGCTGGGAAGGCAAGGGATGC
Gasterosteus -----TGTGACCTATGGGCT TCCCTTTGTCATCCTATGCCTGGGCTCATACTAA--AGAGGCTGGGAGGCAAGGGACGC
Gadus     --T-----GTGACCAATGGGCT TCCCTTTGTCATCCTATGCCTGGGCTCATACTAA--AGAGGCTGGGAGGCAAGGGACGC
Tetraodon -----TGTGACCTATGGACT TCCCTTTGTTATCTCTATGCCTGGA--CTCAAAA--AGGGGCTGGGAGGCAAGGGTTGC
Gallus    -----CCTGTGAGCT TCCCTTTGTCATCCTATGCCTGAG--CG-ATGCTGGAGGCTGGGAGGCTGAAGGGAGGC
Procavia  -----TTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Erinaceus --GGCTACAGTC--TCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT
Tupaia    --GGCTACAGTCTTTCTTCATGTGACTCGTGGACT TCCCTTTGTCATCCTATGCCTGAGAATATATGAA--GGAGGCTGGGAAGGCAAGGGACGT

```

** * ***** :*** * . , * , * , * ***** :

Figura 1.11: Risultato dell'allineamento con clustalW della sequenza del gene *miR-204*. È evidenziato il corrispettivo del miR-204-5p maturo.

Materiali e metodi

Indice

2.1	Il microRNA miR-204-5p	30
2.1.1	Selezione del microRNA da testare	30
2.1.2	Linea cellulare mero-14	30
2.1.3	Protocollo sperimentale	31
	Valutazione della qualità e della quantità dell'RNA	32
2.1.4	Analisi del messaggero	32
	Analisi statistica	33
2.1.5	Analisi della proteina	34
	Quantificazione delle proteine estratte	35
	Western Blot	36
2.2	Analisi comparata in-silico del gene <i>MSLN</i>	37
2.2.1	Allineamenti	37
	Definizione dei confini esone-esone	37
2.2.2	Generazione degli alberi filogenetici	38
	Introduzione dell'outgroup	39
2.2.3	Analisi filogenetica ed evolutiva	40
	Analisi Site Model	41
	Stima del valore Omega per il gene <i>MSLN</i>	42
2.3	Conservazione filogenetica del miR-204	43

2.1 Studio funzionale in vitro dell'interazione tra il microRNA miR-204-5p e la mesotelina

2.1.1 Selezione del microRNA da testare

selmiRNA Consultando diversi lavori di microRNA-profiling tra tessuti sani e tumori di pancreas, ovaio e pleura sono stati annotati tutti i microRNA down-regolati in contesto tumorale, selezionando quelli predetti legare la mesotelina. Si è confrontato se tra questi ve ne fosse almeno uno condiviso, non trovandone alcuno. Si sono quindi cambiati i parametri di ricerca, confrontando tutti i microRNA down-regolati indipendentemente dal legame alla mesotelina: solo il miRNA-204-5p è risultato comunemente de-regolato.

Dal momento che le precedenti predizioni del legame alla mesotelina erano state fatte con tools bioinformatici che si basano su algoritmi che prendono in esame la sola 3'UTR per testare il legame del microRNA, si è utilizzato il software jMiRanda.jar per verificare se anche il miR-204-5p ha la potenzialità di legarsi.

Degli 8 possibili siti di legame riconosciuti, almeno tre avevano energie di legame sufficientemente forti da portare a ritenere che il miR-204-5p si potesse legare alla mesotelina per quanto a livello di sequenza codificante.

2.1.2 Linea cellulare mero-14

L'analisi funzionale è stata condotta in cellule della linea mero-14, che deriva da effusioni pleuriche di un paziente con passata esposizione all'amianto e diagnosticato mesotelioma maligno. Queste cellule sono le medesime su cui era stata condotta la quantificazione dei livelli di messaggero e di proteina come corrispettivo di tessuti di MPM (Melaiu *et al.*, 2011), e sono quindi le più idonee per saggiare gli effetti di trattamenti volti ad alterare l'espressione della mesotelina, dal momento che la proteina è molto espressa.

Il saggio è stato condotto trasfettando le cellule con un microRNA sintetico a doppio filamento che ha la stessa sequenza del miR-204-5p maturo: Syn-hsa-miR-204-5p mi-Script miRNA Mimic (Qiagen). Come controllo negativo si trasfetta un secondo set di cellule con il siRNA All Star Negative Control siRNA (Qiagen), che è validato non

avere omologia per alcun trascritto umano. Per la trasfezione è stato utilizzato l'HiPerfect transfection reagent (Qiagen), una miscela di lipidi neutri e cationici che permette l'internalizzazione cellulare di piccole molecole di RNA a doppio filamento (siRNA e miRNA) e quindi il loro rilascio minimizzando gli effetti aspecifici di off-target.

A seguito di prove preliminari per rifinire il protocollo di trasfezione, si è eseguito l'esperimento di seguito descritto cinque volte in esperimenti indipendenti, gli ultimi due con alcune limitazioni.

2.1.3 Protocollo sperimentale

Semina (1° giorno)

Il giorno precedente alla trasfezione, vengono seminate 80.000 cellule per piastra p60, quantificate con lo Scepter Cell Counter. Il disegno sperimentale prevede l'estrazione di DNA, RNA e proteine dal medesimo campione a 24, 48 e 72 ore dalla trasfezione; per ogni tempo e trattamento si eseguono tre repliche in parallelo, quindi in totale vengono seminate 18 piastre: 3 per il miRNA e 3 per i controlli negativi nei 3 tempi sperimentali.

Trasfezione (2° giorno)

Si controlla che nelle piastre le cellule siano in buona salute ed alla giusta confluenza, quindi si procede con la trasfezione; vengono preparate due mix distinte, una per trattamento, e dispensate in modo omogeneo sulla superficie delle piastre, in modo che l'assunzione sia uniforme.

Estrazione delle 24h (3° giorno)

Per l'estrazione viene utilizzato il Trizol (TRI Reagent[®], Sigma-Aldrich), una miscela di guanidina isotiocianato e fenolo in soluzione monofase che è in grado di portare in soluzione simultaneamente DNA, RNA e proteine a seguito di omogeneizzazione o lisi di cellule e/o tessuti.

La guanidina isotiocianato ed il fenolo sono in grado di denaturare le proteine, una necessità ai fini dell'estrazione poiché porta alla disattivazione delle ribonucleasi, che altrimenti degraderebbero l'RNA. Dal momento che queste sono estremamente comuni, il protocollo di estrazione viene eseguito preferibilmente in ambiente sterile ed utilizzando plastiche RNasi-free.

Per il primo esperimento, condotto in doppietta piuttostochè in tripletta, si è fatta una prova di estrazione a 3h dalla trasfezione, che corrisponde al tempo stimato per l'internalizzazione dei complessi contenenti il microRNA da parte delle cellule, ed avere così un quadro della situazione di partenza. Alla luce dei risultati, si è deciso di seguire un protocollo di estrazione su tre tempi, ossia 24, 48 e 72 ore, e fare per ciascuno tre repliche per condizione sperimentale.

Per gli esperimenti 4 e 5 invece, avendo quasi finito il microRNA, si sono fatte estrazioni solo a 24 e 48 ore dalla trasfezione, sempre in triplette.

Valutazione della qualità e della quantità dell'RNA

La concentrazione dell'RNA estratto è stata valutata allo spettrofotometro.

Il rapporto A_{260}/A_{280} è un indice della contaminazione da proteine, e per essere considerato puro un RNA dovrebbe avere un rapporto pari a 1,8–2. Per la lettura è prevista una diluizione dei campioni di RNA in rapporto 1:35 (68µl H_2O + 2µl RNA estratto) in acqua DEPC.

Dopo la quantificazione, si esegue una elettroforesi su gel di agarosio al 2%; se l'RNA estratto è integro, devono risultare ben evidenti le bande relative agli RNA 18S e 28S, ad una altezza pari a circa 1,9–2Kb e 4–5Kb, rispettivamente.

2.1.4 Analisi dei livelli di espressione del messaggero della Mesotelina

Inizialmente, si valuta se il trattamento con il microRNA ha portato ad alterazioni nel livello di espressione del messaggero della mesotelina; questo effetto non è prova sufficiente né necessaria alla regolazione post-trascrizionale del gene, dal momento che non tutti i microRNA hanno la capacità di indurre la degradazione del messaggero. La prova concreta della regolazione si avrà solo dopo l'analisi dell'espressione della proteina attraverso Western Blot, ma se si osservassero differenze a livello dell'mRNA si avrebbe una prova più forte a sostegno.

Gli RNA estratti sono stati sottoposti a retrotrascrizione, quindi il cDNA è stato analizzato secondo real time RT-PCR con sonde specifiche Taq-man (Applied Biosystem),

che è al momento la metodica più sensibile per valutare quantitativamente l'espressione di cDNA.

La reazione di real time PCR viene avvenire in uno strumento che misura la fluorescenza emessa dalle sonde, e proporzionale ai prodotti di amplificazione. I livelli basali di fluorescenza iniziale non sono rilevabili, dato che i prodotti neo-sintetizzati sono insufficienti, e viene riportato un plateau a livelli minimi di emissione. Successivamente, il numero di copie risultato di amplificazione raddoppia ad ogni ciclo, ed è possibile rilevarne la fluorescenza in forma di curva di amplificazione che segue un andamento esponenziale secondo la funzione teorica $y = 2^x$ (dove x corrisponde al numero di cicli). Dopo la fase esponenziale si ha il raggiungimento di una seconda fase di plateau, in cui uno o più reagenti diventano limitanti e l'amplificazione subisce un arresto.

Per determinare la quantità di stampo iniziale si definisce un valore-soglia (threshold) di fluorescenza, che corrisponde al punto in cui la curva di amplificazione supera il rumore di fondo ed ha inizio la fase esponenziale. Al threshold corrisponde il valore Ct (ciclo-soglia), che rappresenta il numero di cicli necessari per raggiungere il valore-soglia di fluorescenza, misurato in modo automatico dallo strumento ma opportunamente aggiustato manualmente.

Il principio chiave è che a quantità maggiori di template iniziale corrispondono Ct più bassi, mentre basse quantità richiedono un maggior numero di cicli per raggiungere la soglia di fluorescenza. Per ogni reazione e gene analizzato, il Ct impostato per la quantificazione deve essere il medesimo per tutti i campioni. La quantificazione è in termini relativi, misurata come il rapporto di quante volte cambia l'espressione del gene di interesse rispetto ad un riferimento. Da indicazioni preliminari provenienti da uno studio sull'espressione genica nel mesotelioma pleurico maligno, il riferimento è dato da 3 geni housekeeping, *HPRT*, *RPLP0* e *TBP*, analizzati in real time RT-PCR con sonde specifiche. Ogni campione viene analizzato in triplette ai fini di affidabilità statistica.

Analisi statistica

Per rilevare se vi è stato un cambiamento nei profili di espressione del messaggero della mesotelina si deve effettuare una normalizzazione tra repliche sperimentali, calcolando per ciascun campione la Quantity (Q), ossia il prodotto tra l'efficienza di reazione e la differenza tra minCt e la media dei Ct per quel campione:

$$Q = [efficienza * (minCt - media Ct campione)]$$

Si calcola poi la deviazione standard di Q (SD Q) come il prodotto tra la deviazione standard dei Ct per il campione (SD Ct) e la corrispettiva Quantity, moltiplicato per il logaritmo naturale dell'efficienza:

$SD Q = [SD Ct campione * Quantity campione * \ln(efficienza)]$ Questo viene fatto per ciascuno dei tre geni housekeeping e per la mesotelina, e trattandosi di una quantificazione relativa, si normalizzano i valori di Quantity della mesotelina rispetto a quelli dei geni di riferimento: $Q_{campione} / NF_{campione}$. dove NF è il fattore di normalizzazione dato dalla media geometrica tra le Quantity dei geni di riferimento per quel campione. Questo metodo di analisi statistica viene applicato ai risultati di ciascuno dei 5 esperimenti. I valori di mesotelina normalizzati vengono quindi sottoposti ad analisi statistica con il software dedicato StatGraphics. Il disegno sperimentale prevede tre variabili: il trattamento con il microRNA, l'esperimento, ed il tempo di estrazione, e per poterli includere tutti nell'analisi dopo ogni esperimento è stata valutata la variazione dei livelli di espressione del messaggero tra controllo negativo e trattamento nel tempo. In aggiunta, alla fine è stata eseguita una trasformazione (Z-score) sui livelli di mesotelina normalizzata di ogni campione per i cinque esperimenti al fine di eliminare la variabile "esperimento".

I punti di Z-score di tutti i campioni di tutti gli esperimenti possono essere così comparati in un'analisi statistica ANOVA a due variabili: il livello di espressione della mesotelina (controlli negativi vs trattamenti) e tempo di estrazione.

2.1.5 Analisi dei livelli di espressione della proteina msln

Per valutare la concentrazione proteica della mesotelina tra i controlli negativi ed i campioni trattati con il microRNA si esegue un western blot (WB) su estratti proteici dell'esperimento 2.

Alla base del WB vi è una elettroforesi su gel di acrilammide (SDS-page), che ha la finalità di separare le proteine, una volta denaturate, in accordo al peso molecolare. Alla migrazione fa seguito il trasferimento delle proteine su membrana, tipicamente di nitrocellulosa, mediante l'applicazione di una corrente elettrica. Il sandwich (che mette a contatto diretto il gel e la membrana) viene immerso in una soluzione tampone attraverso cui viene fatto passare il campo elettrico, così che le proteine, cariche negativamente

perché sottoposte al trattamento con SDS, si spostino dal gel alla membrana.

Una volta trasferite le proteine su membrana, si fa interagire quest'ultima con l'anticorpo specifico per la proteina che si vuole analizzare.

Per la mesotelina si utilizza come anticorpo primario "Mesothelin-G1", un anticorpo monoclonale di origine murina diretto contro i residui aminoacidici 295–574, in prossimità del C-terminale della proteina umana.

Come riferimento si utilizza l'actina, la cui espressione stabile fa sì che possa essere utilizzata per la normalizzazione dei livelli di espressione, un po' come se fosse l'housekeeping per le proteine. Dal momento che actina e mesotelina hanno peso molecolare simile, bisogna preparare due gel (e due membrane), caricate con i medesimi campioni; una sarà rilevata con l'anticorpo contro la mesotelina e l'altra con l'anticorpo contro l'actina.

Per la rilevazione delle membrane è necessario un anticorpo secondario che interagisca con il primario e che –essendo coniugato con un enzima come una perossidasi o una fosfatasi alcalina, o un fluorocromo– ne permetta la visualizzazione. Entrambi gli anticorpi primari sono stati prelevati da topo, e come anticorpo secondario se ne applica uno generico che reagisce specificamente con anticorpi murini, "Goat anti-mouse IgG-HRP", coniugato alla perossidasi di rafano.

Quantificazione delle proteine estratte

Nonostante il riferimento fornito dall'actina, per ciascun campione è previsto il caricamento della medesima concentrazione di proteine, che devono quindi essere quantificate.

La quantificazione è stata effettuata allo spettrofotometro, sulla base di una retta di taratura costruita su standard a concentrazione proteica nota, così da poter interpolare dalla equazione della retta ottenuta la concentrazione degli estratti proteici:

$$[\text{proteina nel campione } x] = \frac{(\text{valore di assorbanza} - b)/a}{3}$$

Allo spettrofotometro le proteine vengono lette ad una lunghezza d'onda $\lambda = 562\text{nm}$.

Per ogni campione si caricano da protocollo 20µg di proteine in 13µl totali di volume, con la differenza tra estratto proteico e volume finale che viene compensata dall'aggiunta di SDS1%.

Western Blot

Si esegue un protocollo di preparazione per gel discontinui di poliacrilammide, che prevede un gel formato da due matrici a composizione differente, stacking e running: il running gel è quello in cui avviene la corsa elettroforetica vera e propria, mentre lo stacking costituisce la matrice in cui vengono caricati i campioni.

Questo tipo di gel è accompagnato da un sistema basato sul Laemmli buffer (SDS-page), dove le proteine vengono denaturate per mezzo del loro riscaldamento in un buffer contenente SDS (sodio dodecil solfato) ed un agente riducente, il β -mercaptoetanol. L'SDS è un detergente fortemente ionico che conferisce una carica elettrica negativa ogni 2 residui di amminoacido; in questo modo, le proteine assumono conformazione lineare ed un rapporto carica/massa uniforme e proporzionale al peso molecolare, e nella corsa su gel vengono separate in accordo ai pesi molecolari.

Lettura al ChemiDoc MP System (Bio-Rad) Dopo l'incubazione con gli anticorpi ed il trattamento con una soluzione apposita, le membrane vengono rilevate utilizzando il ChemiDoc, uno strumento che monta una CCD camera per la registrazione delle immagini in tempo reale secondo un sistema di rilevamento basato su luce bianca (che permette ad esempio di fotografare il ladder), UV ed altri filtri al LED (per la fluorescenza e la chemiluminescenza).

Sono stati eseguiti due modelli sperimentali di western blot.

Nel primo caso si è analizzata l'espressione delle proteine di tutti e sei i campioni dell'esperimento 2 a 72h dalla trasfezione (tempo ottimale per vedere i risultati di un trattamento con microRNA sulle proteine).

Il secondo modello ha preso in considerazione due tempi di estrazione -48 e 72h dalla trasfezione- scegliendo per ciascuno un controllo ed un trattamento sulla base della concentrazione proteica.

Con il primo modello, si mira ad ottenere una indicazione decisiva sugli effetti del trattamento, se vi sono, mentre con il secondo si vuole mostrare nel tempo la variazione di espressione proteica in relazione al trattamento con il microRNA o in sua assenza.

2.2 Analisi comparata in-silico del gene *MSLN*

2.2.1 Scelta delle sequenze ed allineamenti

Tra tutte le sequenze per gli ortologi della mesotelina depositate sui database NCBI ed Ensembl è stata effettuata una selezione, in modo da scartare quelle con siti non definiti o con un grado di divergenza tale da andare a coprire il segnale evolutivo relativo alla mesotelina umana. Sono state mantenute 16 specie della classe Mammalia, riportate nella tabella 1.1 a pagina 15 della sezione “La mesotelina umana ed i suoi ortologi” nell’introduzione

Per gli allineamenti sono stati utilizzati i seguenti software:

- Clustal W2: <https://www.ebi.ac.uk/Tools/msa/clustalw2/>
- Clustal Omega: <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- Muscle: <https://www.ebi.ac.uk/Tools/msa/muscle/>

Per la sequenza proteica non sono stati necessari accorgimenti, ma dal momento che gli algoritmi su cui si basano i tools bio-informatici hanno difficoltà a gestire come allineare sequenze con lunghi gap e diversità di lunghezza, per il cDNA si è ristretta l’indagine ad una regione compresa tra il sito di inizio della traduzione (ATG) ed il codone di stop. Inizialmente è stata analizzata la sequenza aminoacidica, che presenta il più alto grado di conservazione essendo la proteina sottoposta a vincoli funzionali, quindi sono state comparate le sequenze codificanti e geniche.

Definizione dei confini esone-esone

Si è analizzato se la fase e la l’organizzazione in esoni ed introni è mantenuta a livello comparato; a tal fine, la sequenza completa del cDNA è stata sottoposta al software “Translate tool” disponibile sul sito dell’ExPASy (<http://web.expasy.org/translate/>).

La traduzione in-silico del cDNA restituisce una sequenza con corrispondenza codone-amminoacido, e su questa si effettua manualmente la definizione dei siti di giunzione esone-esone, in particolare per quanto riguarda la proteina, utilizzando come base il tool BLAT fornito dall’UCSC genome browser, che restituisce l’intera sequenza genomica differenziando visivamente gli esoni dagli introni.

Nella traduzione in-silico sono stati segnati i confini di ogni esone, la numerazione, e

la corrispettiva fase, controllando sull'output di BLAT che venissero rispettati i siti accettori e donatori di splicing canonici GT e AG/GC. L'analisi ha permesso inoltre la correzione di alcuni errori di assemblaggio delle sequenze depositate sui database.

Dall'analisi diretta delle sequenze è emerso come all'interno dei Mammiferi sia possibile riconoscere due cluster (blocchi) principali. Nei primati e nei roditori (Euarchontoglires) il gene ha 17 o 16 esoni, in base alla lunghezza della 5'UTR e della 3'UTR, di cui 16 sono quelli codificanti, ma negli altri mammiferi, formalmente i Laurasiatheria, gli esoni codificanti sono 10, a causa di una delezione segmentale che ha rimosso dal gene ancestrale gli esoni 3–8.

Delezioni di tale portata rendono le sequenze ingestibili per qualunque algoritmo, e per questo motivo in un primo momento si è deciso di lavorare con due blocchi distinti, il primo relativo agli Euarchontoglires, ed il secondo relativo ai Laurasiatheria a confronto con *Homo sapiens*.

2.2.2 Generazione degli alberi filogenetici

Gli alberi filogenetici sono stati generati utilizzando gli eseguibili inclusi nel programma “Mega5”, disponibile open-source al sito: <http://www.megasoftware.net/>.

Sia che si esegua l'analisi di una sequenza aminoacidica, sia che si tratti di una sequenza nucleotidica, la modalità è simile, cambiando solamente gli algoritmi, e quindi le impostazioni interne al programma. Per ogni specie, si inserisce la sequenza allineata, così da creare un progetto specifico o per sequenze aminoacidiche o nucleotidiche; Mega5 contiene al proprio interno i programmi clustalW e Muscle per l'allineamento delle sequenze, ma dal momento che questo passaggio è stato fatto in precedenza, non è necessario ripetere l'analisi.

Analisi della sequenze

Si seleziona il menù “Phylogeny”, e si sceglie il tipo di metodo che si desidera utilizzare per la costruzione dell'albero; è stato scelto il Maximum likelihood (WGA*), al momento reputato più affidabile poiché basato su una matrice dei caratteri, eseguendo una analisi con bootstrap (500 repliche). Il bootstrapping è un test statistico per testare i risultati di una indagine filogenetica, e consiste nella creazione di nuovi set di dati (mantenendo l'albero di partenza) campionando quello originario in modo casuale e con

repliche. Un alto valore di bootstrap corrisponde ad un'alta significatività statistica della diramazione, e possono considerarsi robusti nodi con una percentuale almeno del 50%, in accordo alla "majority rule consensus". Se il processo evolutivo si è manifestato in modo uniforme per tutta l'estensione della sequenza considerata, i valori di bootstrap tenderanno ad essere più alti, ma se parti della sequenza hanno subito processi evolutivi differenti, i valori di bootstrap tenderanno ad essere molto bassi.

Il programma genera un albero filogenetico che non assume un modello di evoluzione molecolare (molecular clock), e che risulta quindi senza radice; l'ancestore –ossia la radice– può essere inserita in un punto qualunque sui rami dell'albero. Le lunghezze dei rami rappresentano invece vere distanze evolutive tra specie (il segmento rappresenta l'unità di misura per il numero di sostituzioni rilevate per il ramo specifico), ed accanto a ciascun nodo sono riportati i valori di bootstrap calcolati. Per analizzare una sequenza nucleotidica i passaggi sono i medesimi: si genera il file di input con l'allineamento multiplo e si esegue una analisi di massima verosimiglianza con 500 repliche di bootstrap per sostituzioni nucleotidiche secondo il modello GTR (General Time Reversible). Anche in questo caso, si ottengono alberi senza radice, in cui le lunghezze dei rami corrispondono alla distanza evolutiva tra le specie.

Dagli allineamenti multipli per la sequenza genica completa della mesotelina è emerso un livello di omologia troppo basso ai fini dell'analisi, per cui si è scelto focalizzarsi sulle sole sequenze proteica e codificante (cDNA).

Introduzione dell'outgroup

Per ampliare l'analisi e renderla più solida si è scelto di aggiungere una specie che servisse da outgroup.

Per definizione, in cladistica, con "outgroup" si intende un gruppo monofiletico di organismi (o una singola specie) che non appartiene a quelli in analisi, ma funge da riferimento nella determinazione delle relazioni evolutive tra altri gruppi monofiletici. Nel caso dell'analisi di sequenze, per essere reputata valida, la sequenza deve essere sufficientemente imparentata con quelle in analisi affinché possa esservi comparata, ma meno strettamente di quanto non lo siano le stesse tra loro. Per lo studio è stato scelto l'opossum (*Monodelphis domestica*), per quanto l'omologia di sequenza è molto limitata, in particolare per i primi esoni ed introni (il gene ha 28 esoni, molti dei quali assenti

nei placentati).

E' stato quindi ripetuto l'allineamento a blocchi includendo in entrambi la sequenza dell'opossum, analizzata in relazione a quella umana, quindi sono stati generati gli alberi filogenetici. Un secondo set di alberi filogenetici, necessario per la fase di analisi successiva con PAML, è stato generato utilizzando gli eseguibili nel pacchetto di programmi "Phylip-3.695".

2.2.3 Analisi filogenetica ed evolutiva

Al fine di testare se in particolari linee evolutive la mesotelina è andata incontro a processi di selezione è stato impiegato PAML –Phylogenetic Analysis by Maximum Likelihood– (<http://abacus.gene.ucl.ac.uk/software/paml.html>; Yang), un pacchetto di programmi strutturato sul metodo di maximum likelihood. Questo cerca di quantificare la probabilità che ad una ipotesi H (l'albero filogenetico) corrisponda un certo set di dati D (l'allineamento multiplo). La probabilità viene calcolata come il prodotto delle probabilità che ha ciascun sito dell'allineamento multiplo di presentare un certo carattere se su questo ha agito un processo evolutivo. Tutte le variabili del modello vengono computate per massimizzare il valore di verosimiglianza (likelihood) della ricostruzione filogenetica.

E' stato applicato il programma "codeml", impostando come dati per l'analisi i codoni della sequenza di cDNA. Viene effettuata una comparazione tra modelli sito-specifici, ciascuno basato su differenti assunzioni per quanto riguarda il valore ω ($=dN/dS$), indice utilizzato come parametro per la determinazione di pressione selettiva tra siti.

Vista la notevole divergenza determinata dall'inserimento dell'outgroup nei primi esoni, e la delezione degli esoni 3–8 nei laurasiateri, si è deciso di raggruppare gli allineamenti dei due blocchi in un unico file includente le 17 specie selezionate, considerando tuttavia solo gli esoni dal 9 al 16 (10–17, Poichè l'E1 di uomo rappresenta la 5'UTR), che contengono la porzione di precursore proteico codificante per la mesotelina matura.

Sono state eseguite due analisi, una per i blocchi di esoni E9–10 (279nt); E11 (156nt); E12–13–14 (366nt); E15–16 (273nt), ed una per i singoli esoni, per aumentare la sensibilità al segnale evolutivo.

Analisi Site Model

Il programma codeml è stato eseguito in accordo a cinque modelli evolutivi in una singola computazione:

- M0 (Model 0, NSsite); è un modello base in cui il valore ω è fisso per tutti i rami, ma viene trattato come parametro libero e stimato sulla base delle sequenze fornite.
- M1a (Model 0, NSsite 1); è un modello di quasi-neutralità in cui ω_0 può assumere valori $0 < \omega < 1$ secondo quanto stimato sulla base delle sequenze.
- M2a (Model 0, NSsite 2); è questo un modello di selezione positiva in cui vengono accettati tre parametri ω , $\omega_0 < 1$, $\omega_1 = 1$ ed $\omega_2 > 1$.
- M7 (Model 0, NSsite 7); è un modello di distribuzione beta in cui non sono ammessi siti sottoposti a selezione positiva.
- M8 (Model 0, NSsite 8); è sempre un modello di distribuzione beta, in cui sono ammessi siti sottoposti a selezione positiva.

Omega viene lasciato come parametro da stimare.

Selezionando tutti e 5 i modelli in un singolo run, il programma esegue in automatico due test di rapporto di verosimiglianza (LRT likelihood ratio tests, approfonditi nell'appendice H a pagina 183) per la presenza di selezione positiva in cui un modello nullo –dove $\omega \leq 1$ – viene comparato ad un modello alternativo in cui omega è libero di assumere qualunque valore. Il primo test compara il modello M1a di quasi neutralità con il modello M2a di selezione positiva, ed il secondo compara il modello M7 con M8, con il primo test che viene indicato come più robusto (o meno potente).

Se emerge un qualche dato indicativo al LRT, viene inferita la presenza di pressione selettiva positiva.

A questo risultato viene accompagnato un test per calcolare la probabilità a posteriori che ciascun sito provenga da una particolare classe di siti secondo i modelli NEB (naive empirical Bayes) e BEB (Bayes empirical Bayes). Tra i due, è più affidabile il BEB, che assegna una probabilità a priori ai parametri del modello, e la integra per le loro incertezze, restituendo la posizione e l'identità dell'amminoacido sottoposto a selezione, la media approssimata della distribuzione a posteriori per ω ed il relativo errore standard. In questo modo, è stato possibile inferire se in alcuni esoni sono presenti siti sottoposti a selezione positiva.

Sempre in accordo con l'analisi site model, si è eseguito il modello 3 (M3 discrete; mo-

del=0, NSsite=3), in cui si ammette una pressione selettiva variabile tra siti a 3 parametri liberi nella distribuzione omega: ω_0 , ω_1 , ω_2 . Si esegue quindi il LRT tra i modelli M0, che rappresenta l'ipotesi nulla, ed il modello M3 quale test di variabilità di ω tra siti, piuttostochè come test per la presenza di selezione positiva.

É stato infine eseguito un quarto test per comparare l'ipotesi nulla del modello M8a (che segue sempre una distribuzione beta, ma ha un valore di ω fisso =1, con il modello M8.

Analisi del significato delle sostituzioni aminoacidiche

Per capire se le variazioni significative individuate hanno portato a conseguenze nella struttura e/o funzionalità della proteina, si è utilizzato il programma SIFT

(<http://blocks.fhcrc.org/sift/SIFT.html>), scegliendo un algoritmo basato sull'omologia di sequenza. Il programma restituisce un valore (score) che dipende dagli amminoacidi che compaiono nell'allineamento multiplo e dalle loro proprietà chimico-fisiche, e che rappresenta la probabilità normalizzata che il cambiamento aminoacidico sia tollerato, con un valore soglia di 0,05 al di sotto del quale la sostituzione viene predetta come deleterea per la funzione o la struttura della proteina.

Il possibile stato ancestrale della sequenza è stato ricavato dal tool "Ancestors", Infer Ancestral Sequences (Maximum Likelihood) del programma Mega5, che parendo dall'allineamento multiplo, ricostruisce per ogni nodo dell'albero filogenetico il possibile stato ancestrale dei singoli siti.

Stima del valore Omega per il gene *MSLN*

Per avere indicazioni sul rate evolutivo del gene della mesotelina si sono analizzati i valori di distanza evolutiva a coppie di sequenze per l'intero cDNA e per la porzione di sequenza E9–16.

Per la sequenza ridotta, si è fatto riferimento ai valori di $\omega(dN/dS)$ restituiti dal Site-Model (PAML), mentre per la sequenza completa del cDNA è stato utilizzato Mega5 ("compute pairwise distance").

Lo scopo di questa analisi in-silico è fornire maggiori informazioni sulla mesotelina, tanto a livello di gene quanto di proteina, sfruttando le potenzialità dell'analisi filogenetica comparata per cercare di capire come è evoluta nel tempo la sequenza codificante, se

la filogenia rispetta la filogenesi, o ancora se è andata incontro a cambiamenti importanti in alcune linee filetiche.

2.3 Analisi del livello di conservazione filogenetica dei siti di legame per il miR-204-5p

Per trarre maggiori informazioni sull'effettività del legame del microRNA 204-5p alla mesotelina, si è eseguita una analisi comparata dei siti di legame identificati con jMiRanda.jar per le 16 specie di placentati selezionate per l'analisi filogenetica.

Nei file degli allineamenti multipli a blocchi del cDNA utilizzati per l'analisi filogenetica sono stati evidenziati i possibili siti di nucleazione per valutare se per alcuni di questi vi fosse conservazione a livello filogenetico.

Indice

3.1 Il microRNA miR-204-5p	45
3.1.1 Espressione del messaggero	45
Analisi statistica espressione mRNA	46
3.1.2 Analisi dei livelli di espressione proteica	48
Quantificazione proteica WB mod1	48
Western Blot, modello 1	48
Quantificazione proteica WB modello 2	49
Western Blot, modello 2	49
3.2 Analisi comparata in-silico del gene <i>MSLN</i>	49
3.2.1 Analisi filogenetica ed evolutiva	50
3.3 Conservazione filogenetica del miR-204	55

3.1 Studio funzionale in vitro dell'interazione tra il microRNA miR-204-5p e la mesotelina

3.1.1 Analisi dei livelli di espressione del messaggero della mesotelina

Al fine di valutare se il trattamento delle cellule con il microRNA determina una variazione nei livelli di espressione della mesotelina si è inizialmente analizzato il tra-

scritto del gene. Le tabelle in cui sono riportati i valori ottenuti dalla quantificazione allo spettrofotometro dell'RNA estratto nelle 5 prove sperimentali sono riportate in appendice D.2 a pagina 120.

Analisi statistica dei livelli di espressione del messaggero della mesotelina

Come già descritto nella sezione apposita di materiali e metodi (paragrafo B.4.1 a pagina 88), avendo tre fattori sperimentali da pesare nell'analisi statistica, si è operata una normalizzazione dei valori di mesotelina ottenuti dalle real time RT-PCR in forma di Z-score. In questo modo si riesce a ridurre la variabilità tra prove sperimentali differenti, ed i campioni possono essere analizzati insieme come se facessero tutti parte di una singola sessione in cui le variabili sperimentali si riducono a “tempo” e “trattamento”.

É riportato in figura l'interaction plot dei livelli di espressione della mesotelina tra cellule trasfettate con il microRNA miR-204-5p e cellule di controllo nei tre tempi di estrazione.

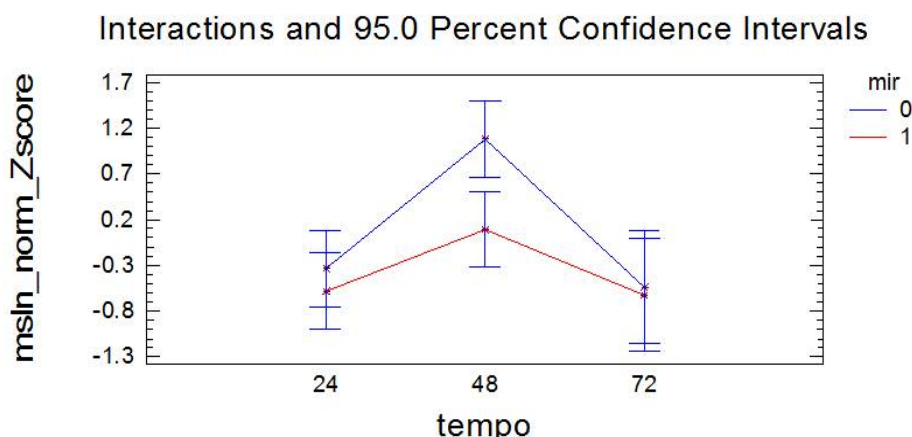


Figura 3.1: Nell'interaction plot sono mostrate le variazioni dei livelli di espressione della mesotelina in funzione del tempo sperimentale di estrazione

L'effetto del trattamento risulta significativo, con un $p\text{-value} < 0,05$, ad indicare come la trasfezione del microRNA determini un abbassamento nei livelli di espressione del messaggero della mesotelina; per aumentare la consistenza della conclusione è richiesta tuttavia anche l'analisi della concentrazione proteica.

Analysis of Variance for msln_norm_Zscore					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-value
Main effect					
A: tempo	15,4028	2	7,7014	16,35	0,00
B: microRNA	2,292	1	2,1920	4,87	0,0322
Interaction					
AB	2,1363	2	1,0681	2,27	0,11
Residual	22,6032	48	0,4709		
TOTAL (corrected)	43,7955	53			

Tabella 3.1: Analisi della varianza per i valori di mesotelina normalizzata

In tutte le repliche sperimentali si è evidenziato un trend nell'espressione del messaggero della mesotelina: basso alle 24h, massimo alle 48h e di nuovo in riduzione alle 72h. Per questo motivo, anche il fattore sperimentale tempo risulta altamente significativo, con un $p\text{-value} < 10^{-5}$. Quello che è importante notare tuttavia è che l'interazione tra i due fattori non è significativa ($p\text{-value}=0,1145$), a dimostrazione di come i risultati dell'esperimento non siano influenzati dai tempi sperimentali, ma solo dal trattamento.

Analysis of Variance for msln_norm_Zscore						
Tempo	Source	Sum of Squares	Df	Mean Square	F-Ratio	P-value
24h	MAIN EFFECT					
	A: miR	0,3257	1	0,3257	0,6	0,4468
	Residual	10,819	20	0,5409		
	TOTAL (Corrected)	11,1447	21			
48h	MAIN EFFECT					
	A: miR	5,4459	1	5,4459	10,22	0,0045
	Residual	10,6543	20	0,5327		
	TOTAL (Corrected)	16,1001	21			
72h	MAIN EFFECT					
	A: miR	0,0179	1	0,0179	0,13	0,7313
	Residual	1,1299	8	0,1412		
	TOTAL (Corrected)	1,1478	9			

Tabella 3.2: Analisi della varianza dei livelli di espressione del messaggero della mesotelina normalizzata, tempo per tempo

Alle 48h si osserva la riduzione maggiore nell'espressione del messaggero della mesotelina nei campioni trattati con il microRNA, un dato compatibile con i tempi di internalizzazione del microRNA, quindi della trascrizione del gene, della maturazione del trascritto ed infine dell'effetto regolatorio del microRNA a livello citoplasmatico.

Nell'appendice J a pagina 195 è possibile trovare l'analisi tempo per tempo ed esperimento per esperimento, in cui i valori "puri" di mesotelina ricavati dalla normalizzazio-

ne per l'espressione genica mediante real-time RT-PCR sono analizzati in relazione al trattamento ed all'esperimento, ed al trattamento ed al tempo, rispettivamente.

3.1.2 Analisi dei livelli di espressione proteica

Quantificazione degli estratti proteici per il modello 1 di Western Blot

Da dati preliminari si aveva indicazione che l'effetto di un trattamento di silenziamento dell'espressione genica si manifesta tendenzialmente dopo 72h dalla trasfezione; nell'appendice D.4.1 a pagina 123 sono riportati i risultati delle letture degli standard, e la retta di taratura interpolata a partire da questi. È stato così possibile determinare le concentrazioni proteiche dei campioni estratti nell'esperimento 2, e quindi i μ l di estratto proteico da caricare per il western blot.

Western Blot, modello 1

Di seguito è riportato il risultato del primo western blot, in cui sono stati analizzati tutti e sei gli estratti proteici dell'esperimento 2 alle 72h:

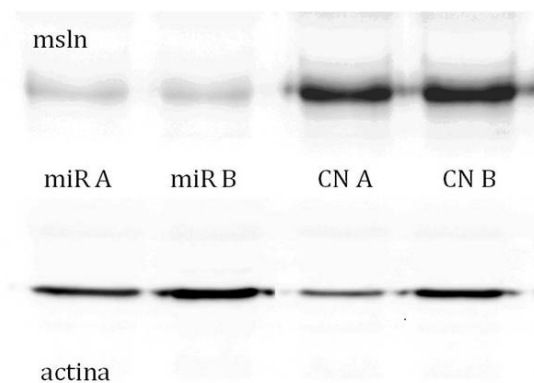


Figura 3.2: Western Blot modello 1 per gli estratti proteici dei campioni alle 72h dell'esperimento 2

Si può apprezzare come l'intensità delle bande per la mesotelina nei campioni trattati con il microRNA è molto più flebile di quella per i controlli negativi; questo indica che il trattamento ha portato ai risultati attesi, e quindi una minore espressione della proteina quando viene fornito il microRNA miR-204-5p.

Quantificazione degli estratti proteici per il modello 2 di Western Blot

Nel secondo modello di western blot sono stati analizzati gli estratti proteici dei campioni dell'esperimento 2 alle 48h ed alle 72h dalla trasfezione; nelle appendici (D.6 a pagina 125) sono riportati i risultati della quantificazione per alcuni campioni scelti in modo casuale tra le repliche. Per la retta di taratura è stata utilizzata quella del modello 1 di western blot.

Western Blot, modello 2

Nel secondo western blot sono stati analizzati un controllo negativo ed un trattamento per le 48h e per le 72h, così da poter confrontare la variazione nei livelli di espressione della proteina.

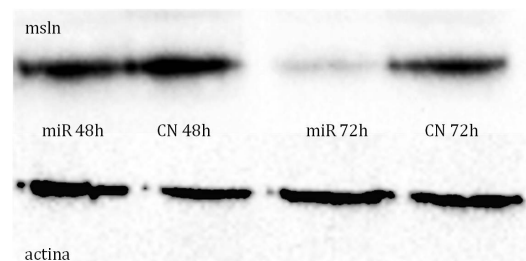


Figura 3.3: Western Blot, modello 2, per gli estratti proteici di campioni alle 48h ed alle 72h dell'esperimento 2

Si può osservare piuttosto chiaramente come il trattamento con il miR-204-5p porti ad un abbassamento nei livelli di espressione della mesotelina alle 72h, quando si riscontra il suo effetto a livello di traduzione.

Queste indicazioni, rafforzate dai dati emersi sull'analisi dei livelli del messaggero, suggeriscono fortemente la capacità del microRNA selezionato di legare la mesotelina e quindi regolarne l'espressione.

3.2 Analisi comparata in-silico del gene *MSLN*

Lo studio comparato in-silico della proteina e del cDNA della mesotelina è mirato a valutare lo stato di conservazione della sequenza codificante e quindi a complementare le indicazioni tratte dall'analisi sul microRNA, che è partita dal presupposto di un

legame tra microRNA e messaggero a livello di sequenza codificante. Nel contempo, si cerca di individuare se nel corso della sua evoluzione il gene è andato ad acquisire –in particolari linee evolutive– caratteristiche peculiari in termini di sostituzioni aminoacidiche o di regioni o siti funzionali che ne hanno possibilmente modificato la funzione o la struttura.

3.2.1 Analisi filogenetica ed evolutiva

Sulla base delle indicazioni emerse di volta in volta, sono stati generati 3 set di allineamenti multipli: uno con tutte le 17 specie, outgroup incluso, uno senza l'outgroup, e l'ultimo senza lo chimpanzee, che essendo troppo simile per certi esoni ad *Homo sapiens* oscurava il segnale evolutivo (ossia un certo grado di divergenza) necessario all'analisi.

Risultati dell'analisi con PAML

I siti (codoni) risultati significativi ai test per l'individuazione di pressione selettiva positiva sono riassunti in tabella 3.3 nella pagina successiva, mentre il LRT è riportato in appendice H a pagina 183.

L'unico sito che in accordo a tutti i modelli e set di allineamenti risulta significativamente sottoposto a selezione positiva è l'amminoacido 482 (217 se si considerano solo gli esoni 9–16). Nei primati l'amminoacido è una leucina, a carattere idrofobico, ma nelle altre linee filetiche si osserva una notevole variabilità per questa posizione. Leucina, alanina, isoleucina e metionina sono tutti amminoacidi idrofobici con catene laterali neutre apolari, e quindi la sostituzione coinvolge amminoacidi con caratteristiche chimico-fisiche analoghe; tuttavia, in *Mus musculus*, *Rattus norvegicus*, *Mustela putorius furo*, *Canis lupus familiaris* e *Monodelphis domestica* si hanno una serina o il triptofano, entrambi idrofili e con catena laterale neutra polare.

a: solo p-values > 0,75

b: p>0.75 secondo il modello M8

c: p>0.75 all'analisi di NEB

nr: non risultato all'analisi

ns: p<0.75

Codone	Probabilità $\omega > 1$			LRT			M3	LRT M0/M3
	cDNA E9–E16	Esone		M1/M2	M7/M8	M8a/M8		
155 K	0,534 b	Esone 12	nr				nr	
172 Y	0,734 b		nr				nr	
185 S	0,591 b		nr				nr	
217 L	0,888 b	Esone 13	0,984*	7,421 ***	6,655 ***	6,347 **	0,997** c	65,562 ***
229 V	0,642 b		0,780 b				0,905 c	
268P	-	Esone 15	0,999**	0,000382	16,5922***	-0,054286	1,00** c	62,481 ***
271V	-		0,899				0,994** c	
282V	-		0,774				0,791 c	
283E	-		1,00**				1,00** c	
289E	-		0,976*				0,985* c	
290R	-		1,00**				1,00** c	
291H	-		0,999**				0,999** c	
292R	-		0,976*				0,985* c	
299L	0,569 b		1,00**				1,00** c	
300R	-		0,965*				0,973* c	
302R	-		1,00**				1,00** c	
304D	-		0,987*				0,993** c	
308T	-		0,999**				0,999** c	
314Q	-		0,999**				0,999** c	
323V	-		0,965*				0,975* c	
326L	-		0,994**				0,995** c	
327S	-		0,999**				0,999** c	
328M	-		0,999**				1,00** c	

Tabella 3.3: In tabella sono riassunti i risultati dell'analisi statistica condotta con PAML per l'individuazione di codoni sottoposti a selezione positiva, analizzando la sequenza del cDNA per gli esoni 9–16, includendo tutte le 17 specie.

Codone	Probabilità $\omega > 1$			LRT			M3	LRT M0/M3
	cDNA E9–E16	Esone		M1/M2	M7/M8	M8a/M8		
172 Y	0,646	Esone 12	nr				0,85 c	
185 S	0,580		nr				0,807 c	
217 L	0,833 b	Esone 13	0,984*	5,513*	6,299**	5,6899*	0,993** c	44,664 ***
229 V	0,541 b		0,809 b				0,846 c	
299 L	nr	Esone 15	0,852 b	0	2,2800	0	ns	68,815 ***

Tabella 3.4: In tabella sono riassunti i risultati dell'analisi statistica condotta con PAML per l'individuazione di codoni sottoposti a selezione positiva, analizzando la sequenza del cDNA per gli esoni 9-16, escludendo l'Opossum dall'analisi.

Codone	Probabilità $\omega > 1$			LRT			M3	LRT M0/M3
	cDNA E9–E16	Esone		M1/M2	M7/M8	M8a/M8		
172 Y	0,753 b	Esone 12	0,843 b	0	1,808	0,75	0,977* c	32,378 ***
185 S	0,620 b		0,851 b				0,977* c	
155 K	0,560 b		ns				0,917 c	
217 L	0,741 b	Esone 13	0,985*	7,639 ***	6,856 ***	6,602 **	0,997** c	65,881 ***
229 V	0,650 b		0,780 b				0,907 c	
268 P	nr	Esone 15	1,00 ** c	0,01372	16,1773***	0,02877	0,999** c	61,604 ***
271 V	nr		0,993** c				0,992** c	
281 H	nr		0,915 c				0,901 c	
282 V	nr		0,788 c				0,772 c	
283 E	nr		1,00** c				1,00** c	
289 E	nr		0,984* c				0,979* c	
290 R	nr		1,00** c				1,00** c	
291 H	nr		0,999** c				0,999** c	
292 R	nr		0,841 c				0,814 c	
299 L	0,592 b		1,00** c				1,00** c	
300 R	nr		0,972* c				0,967* c	
302 R	nr		1,00** c				1,00** c	
304 D	nr		0,992** c				0,989* c	
308 T	nr		0,999** c				0,999** c	
314 Q	nr		0,999** c				0,999** c	
323 V	nr		0,978* c				0,973* c	
326 L	nr		0,995** c				0,994** c	
327 S	nr		0,999** c				0,999** c	
328 M	nr		1,00** c				1,00** c	

Tabella 3.5: In tabella sono riassunti i risultati dell'analisi statistica condotta con PAML per l'individuazione di codoni sottoposti a selezione positiva, analizzando la sequenza del cDNA per gli esoni 9-16, escludendo lo Chimpanzee dall'analisi.

Per valutare se la sostituzione può portare a variazioni nella struttura della proteina, è stato utilizzato il programma SIFT Blink; né il cambiamento da leucina a serina o a triptofano sono predetti comportare variazioni deleteree nella proteina:

Homo sapiens

Substitution at pos 482 from L to S is predicted to be **TOLERATED** with a score of 0,74
(not tolerated: W Y);

Median sequence conservation: 2,36

Sequences represented at this position:43

In questo caso, si è considerato come stato ancestrale dell'amminoacido quello dei primati, e dell'uomo in particolare, motivo per cui l'analisi è stata poi incrementata inferendo lo stato ancestrale dell'amminoacido, ritrovando la seguente condizione (ricostruzione in fig. 3.4 nella pagina successiva):

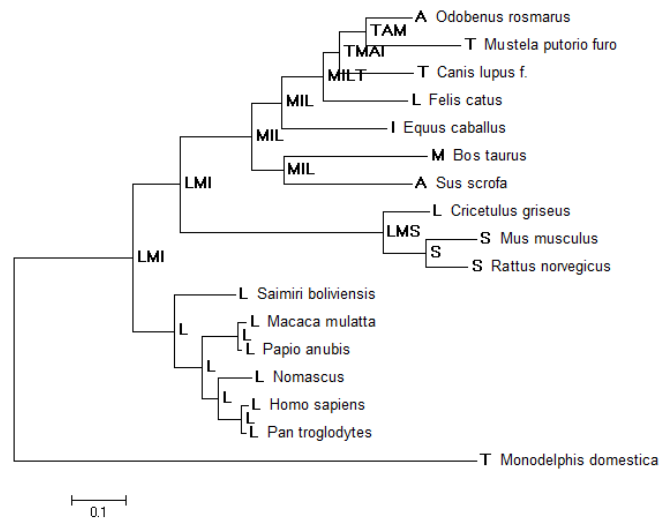


Figura 3.4: Ricostruzione sull'albero filogenetico dello stato ancestrale per l'amminoacido 482

Oltre alla leucina, vengono indicati come possibili amminoacidi ancestrali una metionina o una isoleucina, ed anche in questo caso il SIFT predice le possibili sostituzioni come tollerate.

Lo stesso tipo di analisi è stato condotto per gli altri due siti che, in accordo ad un modello o ad un set di sequenze, sono risultati significativi: l'amminoacido in posizione 494 (229V) e quello in posizione 564 (299L). Anche in questo caso, sostituzioni con amminoacidi dalle proprietà chimico-fisico differenti sono risultati tollerati:

Substitution at pos 494 from V to E is predicted to be **TOLERATED** with a score of 1,00 (not tolerated: W Y F);

Median sequence conservation: 2,43

Sequences represented at this position:42

Substitution at pos 564 from L to F is predicted to be **TOLERATED** with a score of 0,73 (all tolerated);

Median sequence conservation: 2,67

Sequences represented at this position:39

Valori di ω

Si è infine cercato di avere indicazioni sul tasso evolutivo del gene *MSLN* nel suo complesso e limitatamente agli esoni 9–16, nel primo caso utilizzando Mega5 e nel secondo caso i valori ottenuti con PAML secondo i diversi modelli:

	Modello e Tool	ω (dN/dS)
E9–16	PAML: Model 0	0,4400
	PAML: Model 1	0,5381
	PAML: Model 2	0,5381
	PAML: Model 7	0,4634
	PAML: Model 8	0,4828
cDNA	Mega5: average dN/dS	0,6537 (0,219/0,335)
E13	PAML: Model 0	0,5079
	PAML: Model 1	0,6477
	PAML: Model 2	0,7918
	PAML: Model 7	0,5192
	PAML: Model 8	0,6456

Tabella 3.6: Valori medi di ω calcolati con Mega5 per la sequenza completa del cDNA e con PAML per la sequenza del cDNA E9–16

Per quanto i valori non siano esattamente i medesimi (cosa che non sorprende considerando che per ognuno è stato applicato un modello differente) indicano sempre un $\omega < 1$, anche per l'esone 13, che era risultato significativo al LRT.

Questo suggerirebbe che nel corso della sua evoluzione, almeno per il gruppo dei Mammalia, il gene *MSLN* sia stato sottoposto a selezione purificante, e quindi ad una tendenza ad eliminare varianti che avrebbero potuto portare a cambiamenti strutturali e/o funzionali nella proteina codificata.

Per avere possibili indicazioni a supporto di questa conclusione è stato condotto uno Z-test di selezione sui codoni (con bootstrap) con Mega5, mediato sull'intera sequenza codificante per tutte e 17 le specie in analisi, ponendo come ipotesi nulla una selezione neutrale ($dN = dS$) e come ipotesi alternativa la presenza di selezione purificante ($dN < dS$).

Codon-based Test of Purifying Selection for analysis averaging over all sequence pairs:

P-value=0,000

Test statistic (dS - dN)= 8,196

Anche da questo test sembrerebbe quindi che il gene *MSLN* sia stato sottoposto nel corso dell'evoluzione a selezione purificante.

3.3 Analisi del livello di conservazione filogenetica dei siti di legame per il miR-204-5p

Per trarre maggiori informazioni sull'effettività del legame del microRNA miR-204-5p alla mesotelina, si è eseguita una analisi comparata dei siti di legame identificati con jMiRanda.jar per ciascuna delle 16 specie di placentati selezionate per l'analisi filogenetica.

In appendice F a pagina 149 sono riportati gli allineamenti multipli a blocchi in cui sono stati evidenziati tutti i possibili siti di legame del microRNA.

Per quanto riguarda il blocco 1, è emerso un possibile sito di legame condiviso tra tutte le specie, e questo conferisce maggiore supporto alle indicazioni ricavate dagli esperimenti in vitro.

A differenza che nel blocco 1, dove i siti di nucleazione sono concentrati nella prima metà del trascritto, nel blocco 2 è il contrario, e nessuno dei seed-match predetti per *Homo sapiens* trova corrispondenza; se si considera che in questo gruppo filogenetico gli esoni dal 2 all'8 sono assenti per delezione, è più che prevedibile che manchi corrispondenza con i siti del blocco1.

Tuttavia, nei Laurasiatheria sono stati selezionati almeno altri due siti di legame per il micro-RNA, a livello degli esoni 12 e 15.

Indice

4.1	Considerazioni funzionali	57
4.2	Il microRNA 204-5p	61
4.2.1	Considerazioni evolutive sui microRNA	62
4.3	Analisi filogenetica	63
4.4	Cinetica di espressione	65
4.5	Considerazioni conclusive	66

4.1 Considerazioni generali e funzionali sul gene *MSLN*

Nonostante la sua presenza sia stata ormai confermata in tutti i vertebrati, ancora non è chiaro il ruolo biologico svolto dalle proteine codificate dal gene *MSLN*. Poche informazioni si possono reperire anche tentando il confronto con il suo paralogo risultato di duplicazione, il gene *MSLNL* (*Mesotelina-like*), poiché in uomo e gorilla (putativamente anche in chimpanzee) rappresenta uno pseudogene che per selezione rilassata ha accumulato diverse mutazioni deleterie nella regione codificante. In topo entrambi i geni sono attivi, e codificano per un prodotto funzionale; dal database Transcriptome Atlas Database for Mouse Embryo è stato analizzato il profilo di espressione dei due prodotti in embrioni murini:

- *MSLN*: nelle prime fasi di formazione dell’embrione la mesotelina mostra una forte espressione regionale, in particolare a livello di canale alimentare ed intestino, stomaco, retto, e vescica, mentre è solo moderatamente espressa a livello di ventricoli, trachea e diaframma. A 14.5dpc (fig. 4.1) l’espressione rimane fortemente regionale, ma a risultare positiva ai saggi sono le cavità e le loro membrane (diaframma e pericardio, pleura), il sistema escretore, ed i sistemi alimentare e digestivo.

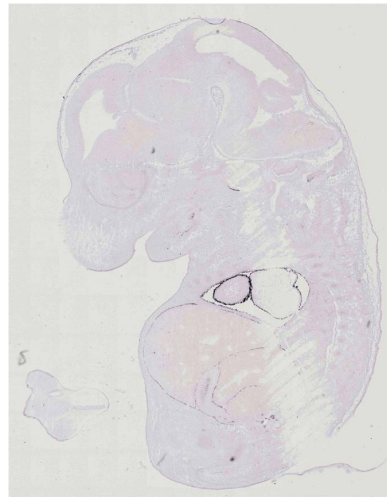


Figura 4.1: Rilevazione immunoistochimica della mesotelina in embrione di topo a 14.5dpc

- *MSLN*: anche per questo gene l’espressione è fortemente regionale, con una spiccata presenza a livello dell’epitelio olfattivo, nonché nei polmoni e nel cuore a 14.5dpc (fig. 4.2 a fronte).

Entrambi i prodotti, identificati come proteine trans-membrana, a livello di struttura terziaria si caratterizzano per una configurazione a super-elica, e mostrano affinità all’otoancorina ed alla stereocilina. Per questo si è proposto che in generale possano legare alla matrice glicoproteica extracellulare la superficie delle cellule cui sono ancorate (Sathyanarayana *et al.*, 2009).

Vi è comunque da sottolineare che non sono stati ancora fatti studi funzionali sulla proteina mslnl, ed anche il coinvolgimento in processi di adesione cellulare è solo inferita sulla base di annotazioni informatiche e computazionali (<http://www.uniprot.org/uniprot/Q8C160>).

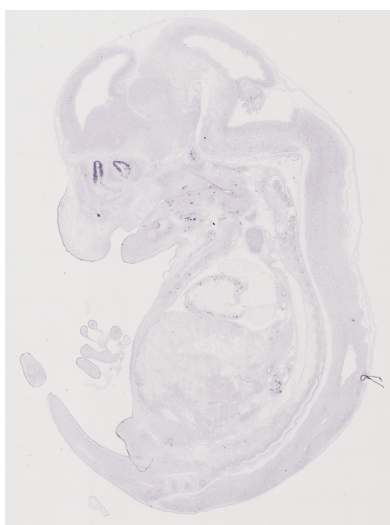


Figura 4.2: Rilevazione immunoistochimica della mesotelina-like in embrione di topo a 14.5dpc

L'assenza del prodotto codificato dal gene *MSLN* nell'uomo è stata interpretata in accordo all'ipotesi "less-is-more": la perdita di questa componente, nonché la degenerazione di alcuni geni interessati nella codifica di recettori olfattivi (Gilad *et al.*, 2003; Matsui *et al.*, 2010) e di geni espressi a livello di epitelio olfattivo come *MOXD2* (Hahn *et al.*, 2007) potrebbe essere associata all'insorgere di nuove caratteristiche fenotipiche nel sistema olfattivo durante l'evoluzione.

Poche sono dunque le informazioni ricavabili dal confronto tra i paraloghi per capire quale sia il suo significato biologico, ed anche dalla consultazione della letteratura la mesotelina risulta nota piuttosto per il coinvolgimento in contesti patologici tumorali, dove risulta sovraespressa. Per quanto riguarda l'uomo, dove si sono concentrate le indagini, molteplici esperimenti hanno mostrato il coinvolgimento della mesotelina nell'incremento del potenziale metastatico ed invasivo delle cellule tumorali attraverso l'attivazione di pathway di modulazione dell'espressione genica (Bharadwaj *et al.*, 2008; Uddalak *et al.*, 2011; Ming-Cheng & Chi-An, 2012; Wang *et al.*, 2012b; Chen *et al.*, 2013) e l'adesione cellulare eterotipica (Scholler *et al.*, 2007; Gubbels *et al.*, 2006). Studi simili sono stati condotti solo in piccola parte in modelli animali (topo e ratto), ma sembra che lo stesso si applichi anche per quelli con sovraespressione o knock-out del gene (Jongsma *et al.*, 2008; Zhang *et al.*, 2011; Servais *et al.*, 2012) o in cui lo sviluppo di mesotelioma pleurico era stato indotto per interazione con fibre di

amianto (Adamson *et al.*, 1993; Marsella *et al.*, 1997; Sandhu *et al.*, 2000; Robinson *et al.*, 2006).

Un dato in particolare ha ispirato il lavoro di tesi: i livelli di espressione del messaggero della mesotelina e della proteina in cellule e tessuti di pleure sane non sono proporzionali ai medesimi riscontrati in tessuti e linee cellulari di mesotelioma pleurico maligno.

Melaiu *et al.* (2011) hanno validato per real-time PCR la sovraespressione del trascritto della mesotelina in tessuti di MPM rispetto a tessuti mesoteliali sani, tuttavia la differenza è debolmente significativa dal punto di vista statistico, mostrando un range di sovrapposizione tra casi e controlli non trascurabile (in fig. 4.3, A). Una migliore definizione è stata conseguita dall'analisi tra cellule Met5A (pleure sane) e mero-14 (MPM), che ha restituito un rapporto di circa 1:2 tra linea cellulare sana e tumorale (in fig. 4.3, B).

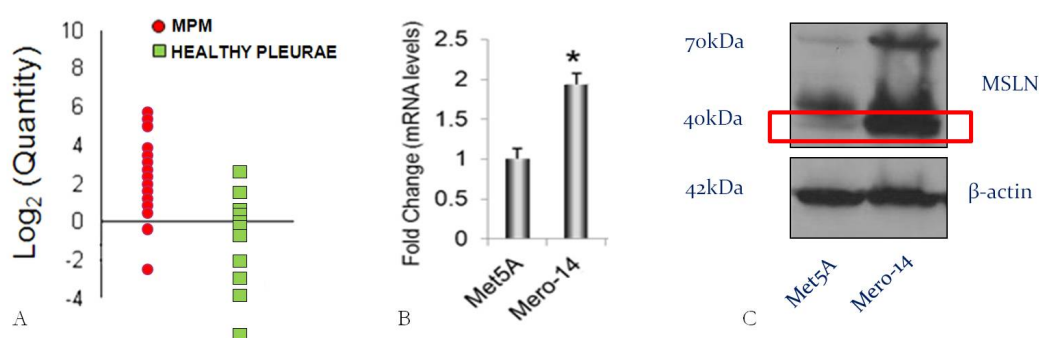


Figura 4.3: Comparazione dei livelli di espressione della mesotelina. A: livelli dell'mRNA tra tessuti sani e di MPM; B: livelli dell'mRNA tra met5A e mero-14; livelli di proteina tra met5A e mero-14.

Se si ammette che la differenza di espressione del trascritto può imputarsi anche alla maggiore accessibilità del promotore in contesto tumorale (Nelson *et al.*, 2011; Tan *et al.*, 2010), e che al livello quasi doppio di messaggero disponibile per la traduzione corrisponde la banda molto marcata nel western blot (riportato in fig. 4.3, C), ci si aspetterebbe un rapporto simile, anche se non strettamente lineare, per quanto riguarda le linee cellulari di tessuto mesoteliale sano. La differenza di espressione proteica è tuttavia molto più rimarcata di quanto non si possa riscontrare per il messaggero, e questo ha portato ad ipotizzare che la discrepanza potrebbe essere dovuta ad un sistema

di regolazione post-trascrizionale nel quadro fisiologico che tiene a freno la concentrazione proteica. In contesto patologico la repressione potrebbe venir meno, e quindi la mesotelina risulterebbe sovraespressa.

I primi regolatori negativi i cui effetti si riscontrano sui quantitativi di proteina presente da prendere in considerazione sono i micro-RNA, ed essendo ormai disponibili in letteratura molti studi su profili di espressione differenziale tra tessuti sani e tumorali ad ampio spettro, si possono sfruttare i dati ricavati come supporto per la ricerca e la validazione di agenti specifici per il contesto di interesse.

4.2 Il microRNA 204-5p

Per quanto riguarda la mesotelina, ancora non sono stati pubblicati lavori che supportino sperimentalmente il ruolo regolatorio svolto dai microRNA. Partendo da uno screening di articoli di microRNA-profiling in studi di tipo caso-controllo, si è selezionato un microRNA down-regolato in tessuti da biopsie di tumore al pancreas, all'ovaio e di mesotelioma pleurico maligno, quindi la sua capacità di legarsi al trascritto della mesotelina è stata confermata con una predizione bio-informatica con il tool jMiRanda.jar. La prima parte del lavoro di tesi è stata incentrata sulla validazione del microRNA miR-204-5p attraverso esperimenti di trasfezione nella linea cellulare mero-14, la medesima su cui erano state condotte le analisi preliminari di espressione del messaggero e della proteina.

Due erano gli elementi più critici a riguardo: il possibile coinvolgimento del microRNA sulla base di indicazioni tratte dalle lavori di microarray, ed il suo legame, predetto alla regione codificante.

È stato riscontrato che quando viene trasfettato il microRNA sintetico si ha un abbassamento tanto nei livelli di espressione del messaggero quanto di proteina sintetizzata, in confronto a cellule di controllo.

A sostegno di queste indicazioni –di per sé molto forti–, si è eseguita una analisi filogenetica comparata su 16 specie di mammifero per i siti di nucleazione del microRNA al messaggero: per quanto riguarda il gruppo tassonomico degli Euarchontoglires, l'analisi delle singole specie con jMiRanda.jar ha rilevato che uno dei diversi seed-match predetti è conservato in tutte le specie, mentre per il gruppo dei Laurasitheria sono stati

individuati due possibili siti di nucleazione differenti, sempre conservati in tutte le linee ed a livello di sequenza codificante.

4.2.1 Considerazioni evolutive sui microRNA

Diversi studi sulla conservazione della regolazione ad opera dei microRNA e sulle sue conseguenze hanno messo in evidenza che siti di legame nella sequenza codificante sono generalmente sottoposti a selezione negativa rispetto a quelli nella 3'UTR (Fang & Rajewsky, 2011). Inoltre, vi sono indicazioni che nelle piante vi sia una più solida conservazione del rapporto tra il microRNA ed il corrispettivo bersaglio, così come degli effetti fenotipici risultanti dall'interazione, mentre questa connotazione è molto meno solida per quanto riguarda i metazoi, rendendo più complessa l'interpretazione dei risultati.

L'analisi non ha permesso di individuare il reale sito di nucleazione, ma si è avuto un ulteriore indizio a supporto della capacità regolatoria del microRNA, poiché l'assenza della regione di match individuata per gli euarcontogliri nei laurasiateri è stata compensata evolutivamente attraverso la selezione di almeno un secondo sito di legame.

Se si accetta questa interpretazione, si ha un forte sostegno al ruolo svolto da questo microRNA nella regolazione negativa in contesto fisiologico dell'espressione della mesotelina.

Tradizionalmente si ritiene che microRNA leganti la regione codificante abbiano una capacità regolatoria inferiore a siti nella 3'UTR, e che la loro presenza (in particolare se conservati filogeneticamente), coadiuvi ed amplifichi le potenzialità repressive di microRNA che legano la 3'UTR.

Non si esclude che questo si applichi anche al miR-204-5p, ma si è riscontrato come la sua trasfezione, che è stato l'unico parametro discriminativo tra il gruppo sperimentale e quello di controllo, determini una regolazione negativa, e si può quindi ipotizzare che in contesti particolari questo da solo sia in grado di portare a differenze nell'espressione del gene target.

4.3 Analisi filogenetica

La presenza di più siti di nucleazione conservati tra i mammiferi è stata il preambolo, e l'ispirazione, per lo studio del gene *MSLN* seguendo un approccio in silico basato sulla comparazione della sequenza della proteina e del cDNA in 17 specie di Mammalia, con l'obiettivo di valutare se nel corso dell'evoluzione la mesotelina ha sviluppato caratteristiche particolari in una o più linee evolutive. Se l'analisi avesse portato a risultati positivi, e quindi all'identificazione di una o più regioni che avessero acquisito caratteristiche distintive in grado di conferire una funzionalità diversa (per quanto non caratterizzata), allora si saprebbe potuto pensare ad un cambiamento della funzione della proteina in una particolare linea evolutiva.

Dalla generazione degli alberi filogenetici per le sequenze proteiche e del cDNA, per intero o limitatamente agli esoni 9–16, si è ricavato che:

- *Monodelphis domestica* (opossum, Marsupialia) rappresenta un outlier rispetto ai placentati, come supportato dalla sua emersione alla base dell'albero e dalla lunghezza del ramo che lo supporta;
- Nel caso del cDNA si distinguono due grandi blocchi principali, uno che include i Primates, ed uno in cui sono inclusi Rodentia e Laurasiateri; nel caso della sequenza proteica completa invece i due grandi blocchi contrappongono da una parte i Laurasiatheria e dall'altra gli Euarchontoglires, come vorrebbe poi la filogenesi di riferimento su base molecolare e paleontologica.
- Tra i Primates il rappresentante dei Platyrrhini (New World Monkey, *Saimiri boliviensis*) emerge alla base, chiaramente distinto dai Catarrhini (Old World Monkey), tra i quali si ha poi una chiara dicotomia tra Hominoidea e Cercopithecoidea.
- Il raggruppamento tassonomico dei Rodentia si presenta solido e costante in tutti gli alberi ricostruiti secondo i diversi parametri, con *Cricetulus griseus* ad emergere per primo. Analizzando la sequenza del cDNA, sia per intero che limitatamente agli esoni 9–16, questi si collocano come sister group dei Laurasiateri, laddove a livello proteico si dispongono come sister group dei Primates.
- Infine, per quanto riguarda i Laurasiatheria, si ha una dicotomia iniziale che oppone gli Artiodactyla ai Carnivora, ma in cui i Perissodactyla (*Equus caballus*) si collocano come

sister group dei secondi. Questo nodo è in molte ricostruzioni supportato da valori di bootstrap non significativi, ma la costanza di questa associazione e valori di bootstrap alti in alcuni alberi portano a pensare che effettivamente per questa specie la storia evolutiva del gene non ricalchi la filogenesi delle specie.

L'analisi filogenetica non ha rilevato accelerazioni evolutive nella porzione di trascritto primario che codifica per la mesotelina. Degli 8 esoni analizzati, solo l'esone 13 ha mostrato per alcuni siti e per il LRT valori indicativi della presenza di selezione positiva, in particolare per l'amminoacido in posizione 482 (RefSeq NP_005814.2 di *Homo sapiens*), che ha restituito valori di $\omega > 1$.

È stato valutato se le variazioni aminoacidiche riscontrate nelle diverse specie a livello di questi siti potessero aver alterato la stabilità e/o la struttura della proteina, ma anche in questo caso non sono emerse conseguenze significative.

Sembrerebbe quindi che nella sua storia evolutiva il gene non sia andato incontro a variazioni sostanziali nella regione codificante, e si può assumere che sia stato invece oggetto di selezione purificante, che ne ha eliminato variazioni che potessero alterare la funzionalità del prodotto codificato.

Il database "OrthoDB" focalizza l'attenzione sull'aspetto evolutivo e comparato di un gene, e quando si ricerca la mesotelina, per tutti i vertebrati viene riportato un rate evolutivo di 1,51 (laddove il gene P53 ha un valore di 1,09, i geni Hox valori < 1 , e la beta globina un valore di 1,02), ben superiore a quanto messo in evidenza dall'analisi mirata e circoscritta eseguita con PAML. OrthoDB presenta un catalogo di geni ortologhi codificanti per proteine ampliato a batteri e metazoi; l'ortologia fa riferimento all'ultimo antenato comune alle specie cui si fa riferimento, ed include così ortologhi a livello di ciascuna radiazione nella filogenia delle specie.

La discrepanza si giustifica tenendo di conto dei parametri e delle scelte che si sono fatte per ricavare il rate evolutivo del gene: sono state analizzate specie della classe Mammalia, e limitatamente alla porzione codificante per la mesotelina, proprio perché oltre un certo livello di divergenza il programma di analisi perde di informazione o non riesce a gestire le differenze. Soprattutto a livello di pesci ed uccelli la divergenza è significativa, fosse solo per la lunghezza del gene e per gli ortologhi riferibili (per lo zebrafish, il gene *BX682557.1-201* –ENSDART00000074557– ha un trascritto lungo 6755bp, e gli si riferiscono come ortologhi tanto il gene *MSLN* quanto il gene *MSLNL* nelle diverse

specie). OrthoDB, per contro, include tutti gli ortologhi catalogati, relativi a 47 specie di Vertebrata, ed il tasso evolutivo è calcolato per ciascun gruppo basandosi sulle identità dei siti in allineamenti multipli proteici per gli ortologhi.

É quindi più che plausibile che a livello di vertebrati il gene presenti un alto rate evolutivo, ma quello che si è messo in evidenza nel presente lavoro è che nei mammiferi placentati il gene si è come stabilizzato, andando di fatto a rappresentare un trascritto altamente conservato.

Ragionevolmente, si può ipotizzare che nei placentati gli aspetti funzionali e regolativi legati alla mesotelina sono rimasti pressoché inalterati fino ad oggi.

4.4 Cinetica di espressione

Dal punto di vista funzionale, un dato interessante emerso dall'analisi dei livelli di espressione del messaggero è che la mesotelina sembra mostrare una espressione dipendente dallo stato proliferativo delle cellule.

Dall'interaction plot in cui i livelli del messaggero della mesotelina normalizzati secondo lo Z-score sono messi in relazione al tempo di estrazione ed al trattamento (fare riferimento alla fig. 3.1 a pagina 46 dei risultati), si può notare come nei controlli negativi vi sia un basso livello di espressione alle 24h, un picco alle 48h ed un nuovo decremento –paragonabile alle 24h– in corrispondenza delle 72h dalla trasfezione e dall'ultimo cambio mezzo. Nell'intervallo tra 24 e 48 ore i nutrienti e lo spazio fisico nelle piastre sono ancora sufficienti a permettere un alto tasso replicativo delle cellule, ma successivamente queste condizioni divengono un fattore limitante (come rivela anche la corsa elettroforetica degli RNA estratti).

Mettendo insieme queste indicazioni, si potrebbe ipotizzare che la cinetica di espressione del gene sia influenzata dallo stato delle cellule, o da una sorta di “quorum sensing”. Questo scenario sarebbe compatibile con la cinetica di espressione di un gene implicato in processi di proliferazione ed adesione cellulare, che è poi quanto suggerito dalla letteratura per quanto riguarda la funzionalità della mesotelina.

4.5 Considerazioni conclusive

Già da anni i principali dogmi della genetica classica sono stati messi a dura prova e quindi corretti per adeguarsi ai nuovi ritrovamenti, e l'analisi comparata dei genomi di specie molto diverse ha permesso lo sviluppo di nuovi modelli per spiegare le variazioni di complessità e le differenze fenotipiche tra linee evolutive. Carroll *et al.* (2008) nell'articolo "Regulating evolution" hanno ad esempio proposto che: «Switches within DNA that govern when and where genes are turned on enable genomes to generate the great diversity of animal forms from very similar sets of genes».

Ad oggi si considera che la diversità tra individui e popolazioni della stessa specie, così come tra specie differenti, è operata da geni molto simili in sequenza e codificanti per proteine che svolgono la medesima attività molecolare, che rispondono ad una regola di "quanto-quando-dove". Le differenze fenotipiche dipendono dalla quantità di proteina, dal quando è espressa (se nell'embrione o nell'adulto, se in risposta a particolari stimoli o regolarmente..) e dal dove, non solo per quanto riguarda tessuti diversi ma anche comparti subcellulari specifici.

É ormai sempre più chiaro come sia il contributo combinato di più componenti regolatorie che agiscono a livelli differenti (trascrizionale e post-trascrizionale), nonché l'interazione di pathway di segnalazione tra cellule e di risposte a stimoli provenienti dall'ambiente esterno a decretare il risultato finale dell'informazione contenuta nei geni. Questo avrebbe permesso lo sviluppo di organismi multicellulari sempre più complessi. Se si accetta dunque che lo sviluppo sia conseguenza della realizzazione di un preciso pattern spazio-temporale di espressione genica, allora è proprio nel contesto dello sviluppo che l'evoluzione dei meccanismi di regolazione genica è più strettamente correlata all'evoluzione degli organismi. All'aumento della complessità biologica è stato riconosciuto un parallelo aumento di complessità nelle famiglie di fattori di trascrizione e nei pathway di regolazione (Chen & Rajewsky, 2007); parallelamente, è stato suggerito che il numero dei geni codificanti per microRNA, la loro espressione e la diversità nei geni loro target sia positivamente correlata alla complessità morfologica osservata nel regno animale (Lee *et al.*, 2007). Sono stati messi in evidenza pathway di regolazione ad opera di microRNA conservati virtualmente inalterati nel corso della filogenesi, mentre la loro diversificazione è stata spesso correlata –anche se non causalmente– con eventi di speciazione.

É questo il messaggio che emerge dal mio lavoro di tesi: anche laddove la sequenza codificante del gene si presenta sottoposta a selezione purificante affinché la funzione biologica del prodotto non venga alterata, le particolarità e possibili differenze di comportamento tra uomo ed altri mammiferi, o tra contesto patologico e fisiologico, dovrebbero attribuirsi a variazioni nel complesso pathway di regolazione del gene, tanto a livello di trascrizione quanto di traduzione.

Conclusioni

Quello presentato è un progetto innovativo per la mesotelina, e in qualche modo anche per le dinamiche dipartimentali, poiché l'analisi di laboratorio basata su tecniche di biologia molecolare e genetica è stata affiancata e sostenuta da uno studio in-silico sulle relazioni evolutive e filogenetiche, portando all'integrazione di due settori differenti e che possono nondimeno trarre vantaggio l'uno dalle potenzialità dell'altro.

Oggetto dello studio è stata la mesotelina, una proteina la cui funzionalità biologica è ancora controversa, e nota piuttosto per la sovraespressione in contesto patologico tumorale, dove è implicata in processi che favoriscono la proliferazione, la sopravvivenza e l'invasività delle cellule che la esprimono.

Si è messo in evidenza come il microRNA 204-5p è capace di esercitare un effetto regolatorio negativo tanto sui livelli di trascritto quanto di proteina, compatibile con la predizione di un appaiamento perfetto tra microRNA e messaggero. Questo dato supporta l'intuizione da cui è scaturita l'indagine secondo cui, in contesto fisiologico, interverrebbe un complesso processo di repressione a carico dell'espressione proteica in cui prendono parte anche microRNA, e che per contro sarebbe alterato in ambito patologico. Questo microRNA ha la peculiarità di legarsi –in accordo alle predizioni bioinformatiche– alla regione codificante del messaggero, circostanza che ha ispirato l'analisi filogenetica comparata del gene, ancora lacunosa in letteratura.

La filogenia della mesotelina ricalca per i Mammalia la filogenesi di riferimento su base molecolare e paleontologica, riconoscendosi chiaramente i cluster di Placentalia e Marsupialia, e tra i primi Euarchontoglires e Laurasiatheria, per quanto poi i rapporti di divergenza interni non sempre corrispondono. In particolare, i Rodentia si associano ai

Primates per la sequenza proteica ed ai Laurasitheria per quella nucleotidica, e si caratterizzano per una lunghezza del ramo di emergenza generalmente maggiore di quanto riscontrato per gli altri placentati.

La porzione codificante per la mesotelina matura non mostra segni di selezione positiva in alcuna linea evolutiva inclusa nell'analisi, e si può ricavare che almeno nei principali gruppi di mammiferi la selezione ha fatto in modo che non venisse alterato significativamente il prodotto codificato.

Dal momento che il tasso evolutivo computato includendo tutti gli orologi riconosciuti nei vertebrati suggerisce variazioni sostanziali nella mesotelina, si può ammettere che questa abbia acquisito nuove caratteristiche prima della divergenza dei placentati, tuttavia le indicazioni ricavate sembrano indicare che al loro interno la funzionalità ancestrale si sia conservata fino ai taxa attuali.

Con predizioni computazionali è stato individuato un sito di nucleazione del microRNA perfettamente conservato nel gruppo degli Euarchontoglires nell'esone 8, ed altri due siti nei Laurasiatheria in corrispondenza degli esoni 6 e 9.

La conservazione filogenetica del microRNA maturo e dei siti di nucleazione suggerisce che la regolazione del miR-204-5p sulla mesotelina sia importante a livello fenotipico, al punto che anche in quelle linee evolutive in cui è assente la regione di legame ancestrale ne è stata evoluta una seconda al fine di mantenere la possibilità di interazione.

Per quanto il coinvolgimento in processi di adesione cellulare e proliferazione sia solo speculativo, ed esperimenti di knock-out indichino che non sia necessaria, il sistema di regolazione molto fine e la selezione purificante che caratterizzano la mesotelina nei Placentalia porta a pensare che la sua funzione non debba essere affatto secondaria.

La combinazione di analisi filogenetico-evolutive e sperimentali di laboratorio si è rivelata uno strumento molto potente, al punto che per il futuro si prospetta di ricorrervi per studi incentrati sulla regione del promotore, per cercare di far luce su elementi che possono partecipare alla tessuto-specificità ed altre modalità di espressione del gene.

Varianti del trascritto primario del gene *MSLN* secondo i diversi database genomici

A seconda del database genomico che si consulta, si riscontrano notevoli differenze nei trascritti primari, o isoforme, che vengono codificati del gene *MSLN*. Sono di seguito riportati i risultati della ricerca nelle banche dati principali.

NCBI

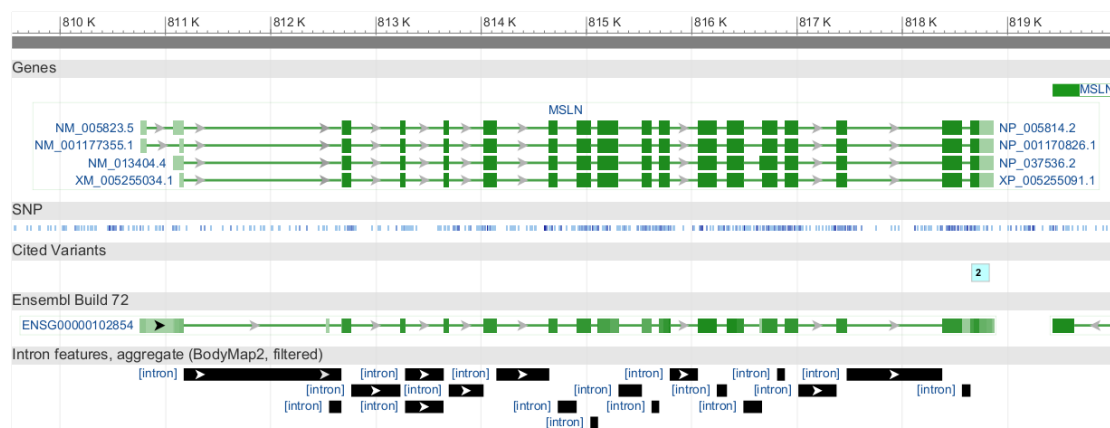


Figura A.1: Isoforme del gene *MSLN* riportate dal database NCBI

Il database dell'NCBI riporta tre varianti di trascrizione:

1. NM_001177355.1 (→ NP_001170826.1), mesothelin isoform 1 preproprotein (2126bp); è considerata la variante 3, che ha una differenza a livello di 5'UTR

rispetto alla variante 1, ma codifica per l'isoforma 1, proprio come la variante 1. (UniProtKB/Swiss-Prot:Q13421)

2. NM_005823.5 (→ NP_005814.2), mesothelin isoform 1 preproprotein (2187bp); denominata come la variante 1, codifica per l'isoforma maggioritaria, ossia la 1. É la sequenza utilizzata come riferimento per l'analisi condotta nel lavoro di tesi. (UniProtKB/Swiss-Pro: Q13421)
3. NM_013404.4 (→ NP_037536.2), mesothelin isoform 2 preproprotein (2154bp); la variante 2 utilizza un sito di splicing alternativo nella regione codificante rispetto alla variante 1, e codifica per l'isoforma 2, caratterizzata da una lunghezza maggiore del trascritto. (UniProtKB/Swiss-Pro:Q13421)

Viene infine riportata una quarta variante, XM_005255034.1 (→ XP_005255091.1), definita variante X1, e deriva da una analisi computazionale automatizzata con il supporto di mRNA ed EST (2041bp).

Ensembl

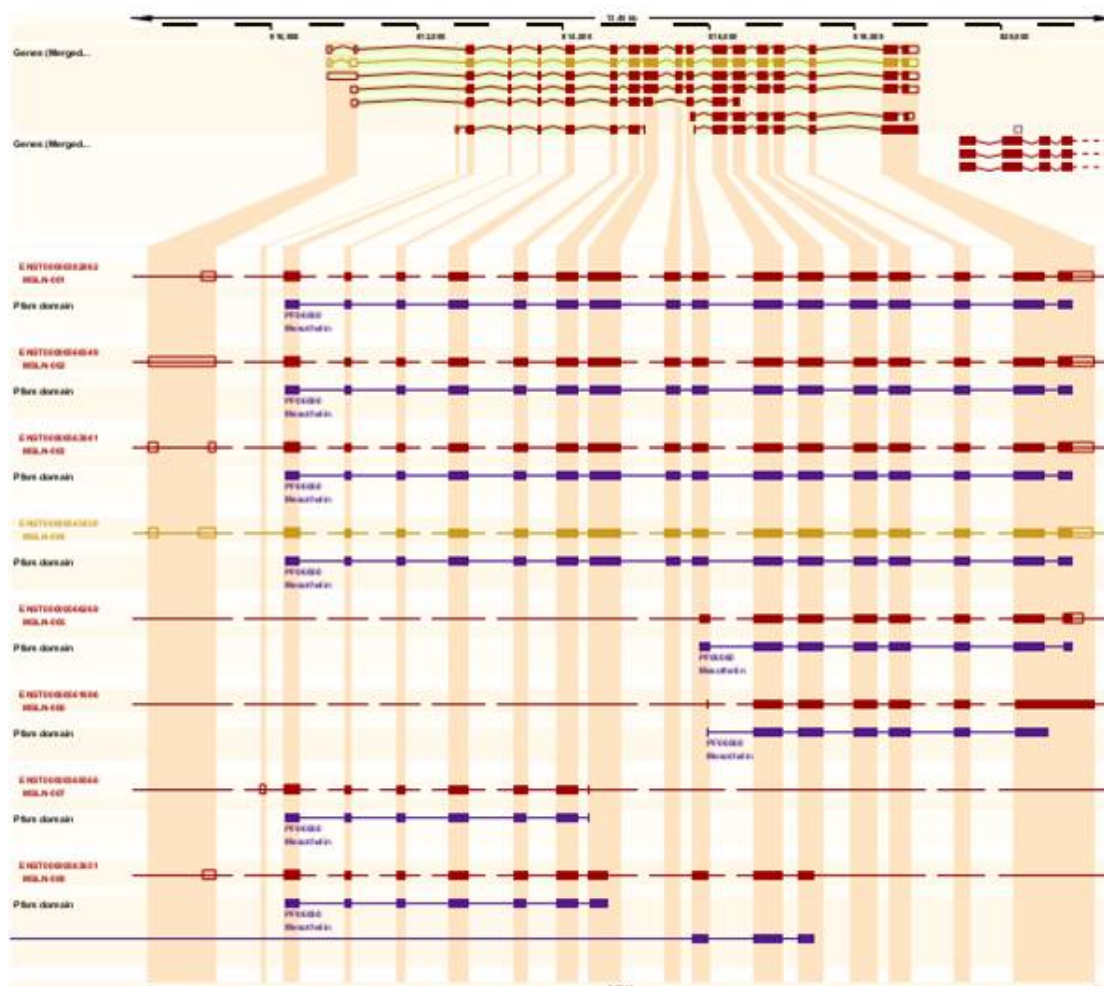


Figura A.2: Isoforme del gene *MSLN* riportate dal database Ensembl

Il database dell'Ensembl riconosce ben 8 varianti di splicing:

1. MSLN-002 (ENST00000566549); 2418bp
2. MSLN-004 (ENST00000545450); 2162bp
3. MSLN-001 (ENST00000382862); 2116bp
4. MSLN-003 (ENST00000563941); 2109bp
5. MSLN-006 (ENST00000561896); 1195bp
6. MSLN-008 (ENST00000563651); 1084bp

7. MSLN-005 (ENST00000566269); 1068bp
8. MSLN-007 (ENST00000569566); 553bp

La MSLN-003 corrisponde alla variante 3 dell'NCBI, MSLN-004 alla variante 1, ed MSLN-001. Le varianti dalla 5 alla 8 segnalano una incompletezza nel CDS a livello di 5'UTR e/o 3'UTR.

UCSC Genome Browser

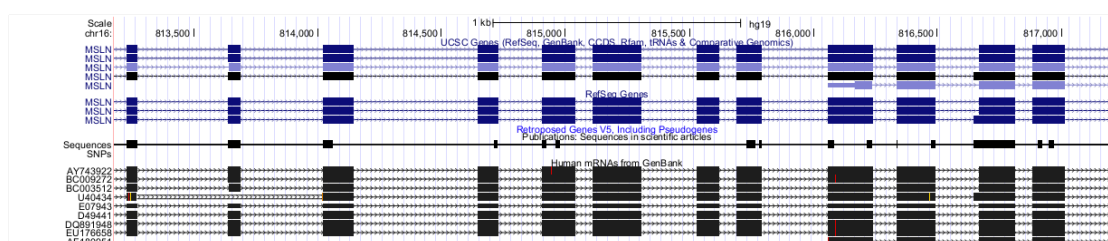


Figura A.3: Isoforme del gene *MSLN* riportate dal database UCSC Genome Browser

Quando si ricerca “MSLN” sul database dell'UCSC, vengono restituite 5 accession tra cui scegliere, che corrispondono a 3 differenti transcript variant:

1. MSLN at chr16:810767-818865 - Homo sapiens mesothelin (MSLN), transcript variant 1, mRNA.
2. MSLN at chr16:816059-818865 - Homo sapiens mesothelin (MSLN), transcript variant 3, mRNA.
3. MSLN at chr16:811073-818865 - Homo sapiens mesothelin (MSLN), transcript variant 2, mRNA.
4. MSLN at chr16:810765-818865 - Homo sapiens mesothelin (MSLN), transcript variant 1, mRNA.
5. MSLN at chr16:810765-818865 - Homo sapiens mesothelin (MSLN), transcript variant 1, mRNA.

Le RefSeq si riducono a tre isoforme:

1. MSLN at chr16:810765-818865 (NM_005823) mesothelin isoform 1 preproprotein
2. MSLN at chr16:810766-818865 (NM_001177355) mesothelin isoform 1 preproprotein
3. MSLN at chr16:811073-818865 (NM_013404) mesothelin isoform 2 preproprotein.

Protocollo dettagliato dello studio funzionale in vitro dell'interazione tra il microRNA miR-204-5p e la mesotelina

B.1 Selezione del microRNA da testare

Non disponendo di ipotesi preliminari, si è condotta una ricerca bibliografica per la scelta del microRNA da testare, con l'unica condizione necessaria che avesse come bersaglio il messaggero del gene *MSLN*. Sono stati valutati più lavori di microRNA-profiling in tessuti di tumore al pancreas, all'ovaio e di mesotelioma pleurico maligno (MPM), tutti caratterizzati dalla sovra-espressione della proteina rispetto ai medesimi tessuti in condizioni fisiologiche.

Tra i microRNA differenzialmente espressi ne sono stati individuati di down-regolati e di up-regolati nel tumore, ma dal momento che la mesotelina è più espressa, sono i primi quelli di interesse.

Sono stati consultati i seguenti lavori:

Pancreas:

- Andrea S. Bauer et al.: Diagnosis of Pancreatic Ductal Adenocarcinoma and Chronic Pancreatitis by Measurement of microRNA Abundance in Blood and Tissue. PLoS ONE 2012 April; 7(4)

Ovaio:

- N Dahiya et al.: MicroRNA expression and identification of putative miRNA targets in ovarian cancer. PLoS One. 2008 Jun 18;3(6):e2436

- MV Iorio et al.: MicroRNA signatures in human ovarian cancer. *Cancer Res.* 2007 Sep 15;67(18):8699-707.
- CH Lee et al.: MicroRNA Profiling of BRCA1/2 Mutation-Carrying and Non-Mutation-Carrying High-Grade Serous Carcinomas of Ovary. *PLoS One.* 2009 Oct 2;4(10):e7314
- GD Miles et al.: Identifying microRNA/mRNA dysregulations in ovarian cancer. *BMC Res Notes.* 2012 Mar 27;5:164.
- EJ Nam et al.: MicroRNA Expression Profiles in Serous Ovarian Carcinoma. *Clin Cancer Res.* 2008 May 1;14(9):2690-5
- SK Wyman et al.: Repertoire of microRNAs in Epithelial Ovarian Cancer as Determined by Next Generation sequencing of Small RNA cDNA Libraries. *PLoS One.* 2009;4(4):e5311
- H Yang et al.: MicroRNA Expression Profiling in Human Ovarian Cancer: miR-214 Induces Cell Survival and Cisplatin Resistance by Targeting PTEN. *Cancer Res.* 2008 Jan 15;68(2):425-33
- L Zhang et al.: Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc Natl Acad Sci U S A.* 2008 May 13;105(19):7004-9

Mesotelioma pleurico maligno:

- C. Croce: unpublished data

Sono stati listati tutti microRNA down-regolati nei tre tipi tumorali, selezionando inizialmente solo quelli predetti legare la mesotelina da tools bioinformatici (TargetScan, miRBase, MiRanda), trovandone 7 per l'MPM, 1 nel pancreas e 7 nell'ovaio, quindi si è confrontato se tra questi microRNA ve ne fosse almeno uno condiviso. Se tre microRNA sono stati trovati a comune tra ovaio e mesotelioma, non c'è stata alcuna corrispondenza per il pancreas.

Si sono quindi cambiati i parametri di ricerca, confrontando tutti i microRNA down-regolati nelle tipologie tumorali, indipendentemente dal legame alla mesotelina, e solo il miRNA-204-5p è risultato comunemente de-regolato.

Dal momento che le precedenti predizioni del legame alla mesotelina erano state fatte con tools bioinformatici che si basano su algoritmi che prendono in esame la sola 3'UTR per testare il legame del microRNA, si è utilizzato il software jmiRanda.jar per verificare se anche il miR-204-5p ha la potenzialità di legarsi.

Questa applicazione si basa su una sequenza di input per il microRNA, nel caso:

>hsa-miR-204-5p (MIMAT0000265): UUCCUUUGUCAUCCUAUGCCU

ed una seconda sequenza di input per il messaggero: >NM_005823

la variante 1 del messaggero della mesotelina, dalla 5'UTR alla 3'UTR.

Il risultato restituito si basa come parametro sul numero di siti che presentano appaia-

mento perfetto alla nucleazione, e quindi sull'energia di legame tra basi complementari.

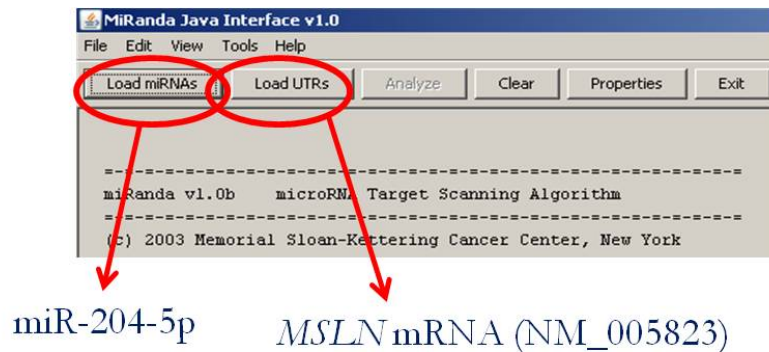


Figura B.1: Interfaccia di jMiRanda.jar

Degli 8 possibili siti di legame riconosciuti, almeno tre avevano energie di legame sufficientemente forti da portare a ritenere che il miR-204-5p si potesse legare efficacemente alla mesotelina, a livello della sequenza codificante.

B.2 Linea cellulare mero-14

L'analisi funzionale è stata condotta in cellule della linea mero-14, che deriva da effusioni pleuriche di un paziente con passata esposizione all'amianto e diagnosticato mesotelioma maligno; queste cellule, con tempo medio di duplicazione di 24h, hanno forma a fuso, ed aderiscono alla superficie delle piastre su cui vengono seminate.

Per la crescita necessitano di un terreno di coltura composto di:

DMEM; 10% siero fetale bovino; 1% pen/strep.

Considerato il tasso replicativo, queste cellule richiedono un cambio di terreno ogni 2-3 giorni, così da eliminare le sostanze di rifiuto e fornire loro nuovo materiale di sostentamento. Le piastre vengono mantenute in incubatore a temperatura costante di 37°C e CO₂ al 5%.

Queste cellule sono le medesime su cui era stata condotta ad un livello preliminare la quantificazione dei livelli di messaggero e di proteina come corrispettivo di tessuti di MPM (Melaiu *et al.*, 2011), e sono quindi le più idonee per saggiare gli effetti di trattamenti volti ad alterare l'espressione della mesotelina, dal momento che la proteina è

molto espressa.

Il saggio è stato condotto trasfettando queste cellule con un microRNA sintetico a doppio filamento che ha la stessa sequenza del miR-204-5p maturo: Syn-hsa-miR-204-5p miScript miRNA Mimic (Qiagen). Come controllo negativo si trasfetta un secondo set di cellule con il siRNA All Star Negative Control siRNA (Qiagen), che è validato non avere omologia per alcun trascritto umano. Per la trasfezione è stato utilizzato l'HiPerFect transfection reagent (Qiagen), una miscela di lipidi neutri e cationici che permette l'internalizzazione cellulare di piccole molecole di RNA a doppio filamento (siRNA e miRNA) e quindi il loro rilascio minimizzando gli effetti aspecifici di off-target.

Per valutare i quantitativi di microRNA e di agente trasfettante più idonei si sono eseguite prove preliminari con il controllo positivo AllStar Hs Cell Death (Qiagen), una miscela di siRNA che interferiscono con l'espressione di geni necessari alla sopravvivenza cellulare. Il controllo negativo è stato trasfettato con il medesimo Negative Control siRNA che verrà utilizzato nelle prove in vitro.

L'esperimento di seguito descritto è stato eseguito cinque volte in esperimenti indipendenti, gli ultimi due con alcune limitazioni.

B.3 Protocollo sperimentale

B.3.1 Semina (1° giorno)

Il giorno precedente alla trasfezione, le cellule devono essere seminate nelle piastre. La quantità da seminare dipende dalla velocità di replicazione delle cellule e deve garantire, il giorno successivo, una confluenza pari al 40–80%. In seguito a prove preliminari, sono state piastrate 80000 cellule per piastra p60 (in fig. B.2 nella pagina successiva): per sapere con precisione quante cellule vengono seminate in ogni pozzetto, vengono staccate tutte quelle cresciute in una piastra p100, e quindi contate con lo Scepter Cell Counter (Millipore), che restituisce il numero di cellule per ml. Basta fare la proporzione tra il numero di cellule totali ed il numero di cellule che si vogliono piastare per sapere quanti ml di cellule dispensare in ogni p60 (5µl di volume complessivo).

Il disegno sperimentale prevede l'estrazione di DNA, RNA e proteine dal medesimo campione a 24, 48 e 72 ore dalla trasfezione; per ogni tempo e trattamento si eseguono tre repliche in parallelo, per un totale di 18 piastre: 3 per il miRNA alle 24 ore, 3 per il

miRNA alle 48 ore e 3 per il miRNA alle 72 ore, e lo stesso per i controlli negativi.

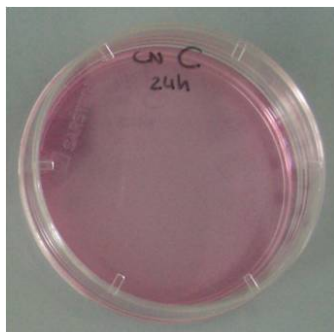


Figura B.2: Esempio di piastra p60 utilizzata per gli esperimenti di trasfezione con il miR-204-5p

B.3.2 Trasfezione (2° giorno)

Si controlla che nelle piastre le cellule siano in buona salute ed alla giusta confluenza, quindi si procede con la trasfezione.

Vengono preparate due mix distinte, una per trattamento, secondo il seguente protocollo:

1. Diluizione del miRNA mimic 204-5p e dell'AllStar Negative Control
I due RNA vengono forniti in soluzioni 20 μ M che vengono diluite di un fattore 1:10 così da avere una aliquota del volume opportuno 2 μ M.
2. Mix di trasfezione (per campione, vol finale= 4 μ l):
 - 100 μ l di terreno DMEM serum free; si tratta di una aliquota del terreno fornito dalla casa produttrice e che viene messa da parte in una falcon al momento dell'apertura, prima che vengano aggiunti il siero fetale bovino e la pen/strep.
 - Aggiungere 10 μ l di miR-204-5p o di AllStar Negative Control 2 μ M -in base alla mix di trasfezione- così da avere per campione una concentrazione finale di 5nM per piastra.
 - Aggiungere 20 μ l di HiPerFect.
 - Vortexare le due mix.
 - Aspettare 10min a T ambiente, in modo che si formino i complessi.
3. Nei 10 minuti di attesa si cambia il terreno alle cellule, aggiungendovi 3870 μ l di terreno DMEM completo.
4. Per ogni piastra distribuire uniformemente 130 μ l di mix di trasfezione, a seconda del trattamento.

5. Mettere le piastre in incubatore.

B.3.3 Estrazione delle 24h (3° giorno)

Per l'estrazione viene utilizzato il Trizol (TRI Reagent[®], Sigma-Aldrich), una miscela di guanidina isotiocianato e fenolo in soluzione monofase che è in grado di portare in soluzione simultaneamente DNA, RNA e proteine a seguito di lisi cellulare.

La guanidina isotiocianato ed il fenolo sono in grado di denaturare le proteine, una necessità ai fini dell'estrazione poiché porta alla disattivazione delle ribonucleasi (RNasi), che altrimenti degraderebbero l'RNA. Dal momento che queste sono estremamente comuni, il protocollo di estrazione viene eseguito preferibilmente in ambiente sterile ed utilizzando plastiche RNasi-free.

Protocollo di estrazione:

I) Lisi cellulare; per piastra

1. Prendere le piastre e togliervi il terreno di coltura;
2. Coprire in modo uniforme la superficie con 750µl di Trizol, quindi spipettare con decisione affinché per azione meccanica le cellule si stacchino e vengano lisate;
3. Prelevare tutto il volume di liquido e trasferirlo in una eppendorf da 2ml;
4. Aspettare 5min a T ambiente, durante i quali si ha la dissociazione completa dei complessi nucleo proteici nel lisato.

II) Separazione delle fasi; per eppendorf

1. Aggiungere 200µl di cloroformio ed agitare vigorosamente;
2. Aspettare 10min a T ambiente;
3. Centrifugare a 12000xg a 4°C per 15min;

L'aggiunta di cloroformio (per l'eliminazione delle tracce di etanolo) ad una soluzione a base fenolica (Trizol) determina una separazione di fasi, come mostrato in fig. B.3 a fronte.

La fase superiore, acquosa ed incolore, contiene l'RNA totale; quella intermedia, all'interfase e più densa, contiene il DNA ed i lipidi di membrana; infine, vi è una fase organica, dal colore rosa tipico del Trizol, che contiene le proteine e rimanenti precipitati cellulari.



Figura B.3: Fotografia della separazione di fasi

III) Estrazione dell'RNA totale

1. Prelevare con attenzione la fase acquosa contenente l'RNA e trasferirla in una eppendorf pulita da 1,5ml.
2. Aggiungere 500 μ l di isopropanolo, che serve per la precipitazione dell'RNA, dal momento che costituisce un solvente moderatamente apolare mentre l'RNA, come tutti gli acidi nucleici, è una molecola polare (carica negativamente grazie ai gruppi fosfato PO_3^{2-}) e perciò perfettamente solubile in acqua e poco solubile in solventi organici. Il cambiamento di solvente da acqua a isopropanolo, combinato con la centrifugazione, determina la precipitazione dell'RNA. Si aggiunge poi 0,5 μ l di glicerolo, che facilita la sedimentazione di molecole di RNA molto piccole, come i microRNA:
3. Aspettare 10min a T ambiente.
4. Centrifugare a 12000xg a 4°C per 10min; si ha così la precipitazione di un pellet di colore chiaro sul fondo della eppendorf.
5. Svuotare il sovrantante facendo attenzione a non perdere il pellet.
6. Aggiungere 950 μ l di EtOH al 75% (in acqua DEPC) ed agitare in modo che il pellet si stacchi.
7. Centrifugare a 7600xg a 4°C per 5min.
8. Prelevare con una pipetta tutto il sovrantante, senza toccare il pellet.
9. Lasciare asciugare il pellet di RNA all'aria per il tempo necessario, solitamente 10–15min, controllando però che non secchi troppo, o si potrebbe degradare l'RNA.

10. Dissolvere il pellet in 50µl di acqua DEPC ed incubare in bagnetto a 58°C per 10min.
11. Conservare l'RNA a -80°C.

IV) Estrazione delle proteine

1. Aggiungere alla eppendorf con le fasi di DNA e proteine 300µl di EtOH 100% ed agitare vigorosamente; l'etanolo assoluto determina la precipitazione delle molecole di acidi nucleici, quindi del DNA.
2. Aspettare per 2min a T ambiente.
3. Centrifugare a 2000xg a 4°C per 5min.
4. Prelevare il sovrnatante, che contiene le proteine e le altre componenti organiche in soluzione, e metterlo in una eppendorf pulita da 2ml; per i miei fini sperimentali il DNA non è richiesto, per quanto si possa estrarre, quindi viene buttato.
5. Aggiungere 950µl di isopropanolo ed agitare vigorosamente; l'isopropanolo permette in questo caso la precipitazione delle proteine.
6. Aspettare 10min a T ambiente.
7. Centrifugare a 12000xg per 10min.
8. Scartare il sovrnatante, facendo attenzione a non perdere il pellet.
9. Lavare il pellet con 950µl di guanidina idrocloride 0,3M in EtOH al 95%; le proteine possono essere lasciate in guanidina a -20°C oppure si può portare l'estrazione fino alla fine:
10. Lasciare le proteine per 20min a T ambiente in guanidina idrocloride; centrifugare a 7500xg per 5min a 4°C.
11. Ripetere il lavaggio con guanidina idrocloride al punto precedente per un totale di 3 passaggi.
12. Lavare il pellet in 1ml di EtOH assoluto e vortexare.
13. Centrifugare a 7500xg per 5min a 4°C.
14. Rimuovere l'etanolo e lasciare asciugare il pellet all'aria per il tempo necessario, solitamente 10–15min.
15. Disciogliere il pellet in 300–500µl di SDS1% (10% di SDS10% + 90% H₂O) e spipettare bene per farlo risospendere; se il pellet è molto compatto, si può lasciare la eppendorf in un bagnetto a 50°C fino a che non si dissolve.
16. Centrifugare a 10000xg per 10min a 4°C per far precipitare il materiale insolubile.
17. Trasferire il sovrnatante contenente le proteine in una eppendorf pulita da 1,5ml, quindi conservare a -80°C.

Estrazione delle 48h (4° giorno)

Seguire il protocollo delle 24h.

Estrazione delle 72h (5° giorno)

Seguire il protocollo delle 24h.

Per il primo esperimento, condotto in doppietta piuttostochè in tripletta, si è fatta una prova di estrazione a 3h dalla trasfezione, che corrisponde al tempo stimato per l'internalizzazione dei complessi contenenti il microRNA da parte delle cellule, ed avere così un quadro della situazione di partenza. Alla luce dei risultati, si è deciso di seguire un protocollo di estrazione su tre tempi, ossia 24, 48 e 72 ore, e fare per ciascuno tre repliche per condizione sperimentale.

Per gli esperimenti 4 e 5 invece, avendo quasi finito il microRNA, si sono fatte estrazioni solo a 24 e 48 ore dalla trasfezione, sempre in triplette.

B.4 Valutazione della qualità e della quantità dell'RNA

Dopo il protocollo di estrazione, la concentrazione dell'RNA è stata valutata allo spettrofotometro.

Il rapporto A260/A280 è un indice della contaminazione da proteine, e per essere considerato puro un RNA dovrebbe avere un rapporto pari a 1,8–2. Per la lettura è prevista una diluizione dei campioni di RNA in rapporto 1:35 (68µl H₂O + 2µl RNA estratto) in acqua DEPC. Lo strumento utilizzato (Gene Quant[®], Amersham Bioscience) fornisce automaticamente la concentrazione dell'RNA, una volta impostato il fattore di diluizione; inoltre, viene indicato il rapporto A260/A230, un secondo indice di purezza per la contaminazione da sali o residui di fenolo o carboidrati nell'estratto.

Dopo la quantificazione, si esegue una elettroforesi su gel di agarosio al 2%; se l'RNA estratto è integro, devono risultare ben evidenti le bande relative agli RNA 18S e 28S, ad una altezza pari a circa 1,9–2Kb e 4–5Kb, rispettivamente.

B.4.1 Analisi dei livelli di espressione del messaggero

Inizialmente, si valuta se il trattamento con il microRNA ha portato ad alterazioni nel livello di espressione del messaggero della mesotelina; questo effetto non è prova sufficiente né necessaria alla regolazione post-trascrizionale del gene, dal momento che

non tutti i microRNA hanno la capacità di indurre la degradazione del messaggero. La prova concreta della regolazione si avrà solo dopo l'analisi dell'espressione della proteina attraverso Western Blot, ma se si osservassero differenze a livello dell'mRNA si avrebbe una prova più forte a sostegno.

Retrotrascrizione

La retrotrascrizione permette di produrre una molecola di cDNA (DNA complementare) a partire da una molecola stampo di RNA. La reazione è catalizzata dalla trascrittasi inversa, un enzima polifunzionale con attività di DNA polimerasi RNA-dipendente ed RNasiH, in modo da garantire l'idrolisi degli ibridi DNA:RNA.

A partire da una miscela di primers differenti, l'enzima sintetizza un filamento di DNA complementare all'RNA templatato, con formazione dell'elica ibrida DNA:RNA e quindi la degradazione del filamento di RNA. Il DNA sintetizzato agisce invece da stampo per la sintesi di una doppia elica di DNA.

Per la retrotrascrizione viene utilizzato il kit iScriptTM cDNA Synthesis kit (bio-rad), che fornisce una trascrittasi inversa RNasiH⁺ ed una miscela di primer oligo(dT), che si appaiano alla coda poli(A) del messaggeri, e di random-esameri, che permettono l'appaiamento ad una grande varietà di templati differenti. Per tanto, un'unica reazione di retrotrascrizione permette la sintesi dei cDNA corrispondenti alla quasi totalità degli RNA contenuti nel campione.

Per la reazione è richiesta una quantità di RNA templatato totale pari a 1µg, e viene condotta in un volume totale di 20µl, così suddivisi:

- . 5x iScript reaction mix: 4µl
- . iScript reverse transcriptase: 1µl
- . H₂O RNasi-free (da kit): xµl a volume finale= 20µl
- . RNA del campione: xµl, pari a 1µg

Per calcolare la quantità in µl di RNA da aggiungere, si prende la concentrazione risultante dalla lettura allo spettrofotometro (ng/µl) e si fa la proporzione:

$$[RNA] : 1\mu l = 1\mu g : x\mu l$$

Quindi, da 15µl si sottrae il valore x calcolato, e si ottiene il volume di acqua RNasi-free da aggiungere.

La mix viene quindi sottoposta a termociclizzazione secondo il protocollo:

- . 25°C per 5min
 - . 42°C per 30min
 - . 85°C per 5min
 - . mantenimento eventuale a 4°C fino a che non vengono prelevati i campioni.
- I prodotti di retrotrascrizione si conservano a -20°C.

Real Time RT-PCR

La real time PCR è al momento la metodica più sensibile per valutare i livelli di espressione di RNA sottoposti a retrotrascrizione. Come la PCR tradizionale si basa sull'utilizzo di una coppia di primers specifica per la formazione dell'amplificato di interesse, ma la sintesi delle copie è seguita nella sua crescita esponenziale ciclo per ciclo.

I due metodi più comuni per la visualizzazione dei prodotti di amplificazione prevedono l'utilizzo di:

- coloranti fluorescenti aspecifici che si intercalano agli acidi nucleici (ad esempio, EvaTM Green);
- sonde specifiche marcate con un fluorescente che permette la visualizzazione del prodotto di PCR solo dopo l'appaiamento della sonda al target, in una regione univoca e specifica per il messaggero di interesse (ad esempio, le sonde TaqMan).

La prima metodica ha il vantaggio di essere più semplice e meno costosa, la seconda è invece più sensibile e specifica.

Per l'analisi di espressione del messaggero del gene della mesotelina si è scelto di utilizzare una sonda (Applied Biosystem) che come i primers forniti è disegnata per appaiarsi in modo specifico ed univoco al messaggero MSLN, garantendo una efficienza di reazione pari al 100%.

Una sonda è una sequenza oligonucleotidica (20nt circa) che presenta coniugate al 5' ed al 3' due molecole, rispettivamente un fluorocromo emettitore (R), che agisce da reporter, ed un quencher (Q). Quando la sonda è ibridata o libera nel mezzo ha conformazione tale da non emettere fluorescenza, poiché si instaura un fenomeno fisico di risonanza (FRET, Fluorescence Resonance Energy Transfer) per cui il fotone emesso da R viene catturato da Q. Nel momento in cui la prossimità fisica necessaria al FRET viene meno

(ossia quando la sonda viene degradata, ed R e Q sono liberi in soluzione) si ha emissione di segnale rilevabile come fluorescenza sotto opportuna stimolazione UV.

Perché questo sia possibile, deve essere utilizzata una polimerasi particolare, dotata di una subunità con attività esonucleasica 5' → 3'; in questo modo, quando trova la sonda ibridata, la polimerasi non si arresta ma la degrada, permettendo la prosecuzione della reazione e la liberazione della fluorescenza.

Ad ogni ciclo, ed in funzione del numero di target-templato, si ha una sonda che si degrada ed un fluorocromo che emette, con l'intensità di fluorescenza emessa che rappresenta quindi una misura direttamente proporzionale alla quantità di prodotto formato. La master mix utilizzata, 5x HOT FIREPol® EvaGreen® qPCR Mix Plus, no ROX (Solis BioDyne, Medibena), contiene la HOTFIREPol® DNA Polimerasi, una DNA polimerasi termostabile che si attiva dopo 15min a 95°C, così da impedire amplificazioni aspecifiche a temperatura ambiente. Sono già posti in soluzione, inoltre, dNTPs ultra-puri, MgCl₂ ed il dye EvaGreen® (un colorante che si lega al DNA senza inibire o ostacolare la reazione di amplificazione).

La reazione di real time PCR viene performata in uno strumento in grado di misurare la fluorescenza emessa dai prodotti di amplificazione (iCycler Thermal Cycler®, Bio-Rad). Inizialmente i livelli basali di fluorescenza non sono rilevabili dallo strumento, dato che i prodotti neo-sintetizzati sono insufficienti, e viene riportato un plateau a livelli minimi di emissione. Successivamente, il numero di copie risultato di amplificazione raddoppia ad ogni ciclo, ed è possibile rilevarne la fluorescenza in forma di curva di amplificazione che segue un andamento esponenziale secondo la funzione teorica $y = 2^x$ (dove x corrisponde al numero di cicli). Dopo la fase esponenziale si ha il raggiungimento di una seconda fase di plateau, in cui uno o più reagenti diventano limitanti e l'amplificazione subisce un arresto (raffigurato in fig B.4 nella pagina successiva).

Per determinare la quantità di stampo iniziale si definisce un valore-soglia (threshold) di fluorescenza, che corrisponde al punto in cui la curva di amplificazione supera il rumore di fondo ed ha inizio la fase esponenziale. Al threshold corrisponde il valore Ct (ciclo-soglia), che rappresenta il numero di cicli necessari per raggiungere il valore-soglia di fluorescenza, misurato in modo automatico dallo strumento ma opportunamente aggiustato manualmente.

Il principio chiave è che a quantità maggiori di templato iniziale corrispondono Ct più bassi, mentre basse quantità richiedono un maggior numero di cicli per raggiungere la

soglia di fluorescenza. Per ogni reazione e gene analizzato, il Ct impostato per la quantificazione deve essere il medesimo per tutti i campioni.

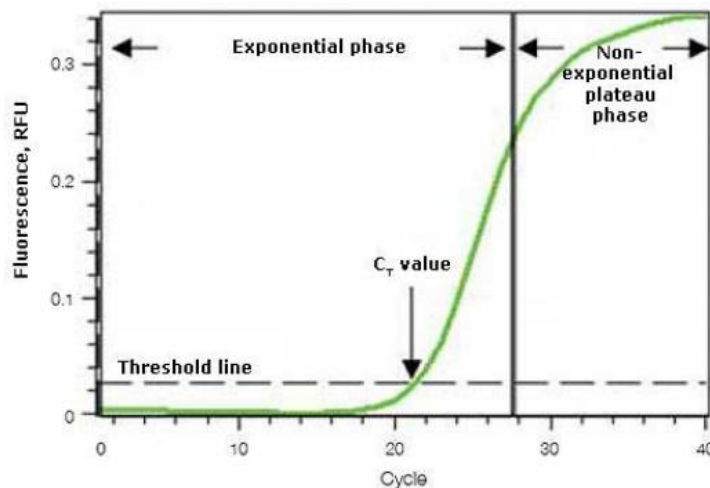


Figura B.4: Rappresentazione di una tipica curva di amplificazione di real time PCR

La quantificazione è in termini relativi, misurata come il rapporto di quante volte cambia l'espressione del gene di interesse rispetto ad un riferimento. Da indicazioni preliminari provenienti da uno studio sull'espressione genica nel mesotelioma pleurico maligno, il riferimento è dato da 3 geni housekeeping, *HPRT*, *RPLP0* e *TBP*, analizzati in real time RT-PCR con sonde specifiche:

- HPRT: HS01003267_m1
- RPLP0: HS99999902_m1
- TBP: HS00427620_m1

Sonda per la mesotelina: HS00245879_m1

Protocollo di qPCR:

Ogni campione viene analizzato in triplette per una maggiore affidabilità statistica.

Reagenti	Vol x1 (µl)
H ₂ O sterile	12
Sonda	1
Buffer FAM qPCR Mix Plus	5
cDNA (1:10)	2

(a) Protocollo di qPCR

	T (°C)	tempo
	95°	15min
40 cicli	95°	15sec
	60°	30sec

(b) Profilo di termociclistizzazione

Tabella B.1: Protocollo di real time qPCR.

Analisi statistica

Il primo obiettivo è valutare se a seguito del trattamento con il microRNA miR-204-5p si riesce a rilevare un cambiamento nei profili di espressione del messaggero della mesotelina, utilizzando per la normalizzazione tra repliche sperimentali il seguente protocollo, che viene abitualmente utilizzato per l'analisi statistica di lavori di gene expression:

- Per ciascun campione sperimentale si segna il valore Ct di emergenza del relativo gene analizzato;
- Calcolare la media delle tre repliche per ciascun campione e la corrispettiva deviazione standard (SD). Affinché la tripletta sia valida, la SD deve essere minore o uguale a 0,4 se il Ct di emergenza è minore di 30, o minore o uguale a 0,6 se il Ct è maggiore di 30;
- Evidenziare il valore minimo tra le medie dei Ct (minCt);
- Per ciascun campione calcolare la Quantity (Q), ossia il prodotto tra l'efficienza di reazione e la differenza tra minCt e la media dei Ct per quel campione:

$$Q = [efficienza * (minCt - media Ct campione)]$$

- Per ogni campione calcolare la deviazione standard di Q (SD Q) come il prodotto tra la deviazione standard dei Ct per il campione (SD Ct) e la corrispettiva Quantity, moltiplicato per il logaritmo naturale dell'efficienza:

$$SD Q = [SD Ct campione * Quantity campione * \ln(efficienza)]$$

I valori indicati vengono calcolati per ciascuno dei tre geni housekeeping e per la mesotelina.

Trattandosi di una quantificazione relativa, è necessario normalizzare i valori di Quantity della mesotelina rispetto a quelli dei geni di riferimento:

- Il fattore di normalizzazione per ciascun campione corrisponde alla media geometrica tra le Quantity dei geni di riferimento per il medesimo. Ad esempio, per il campione x :

$$NF = \sqrt[n]{Ref1_x * Ref2_x * \dots * Refn_x}$$

dove n è il numero di geni di riferimento selezionati, mentre Ref indica la Quantity di ogni gene di riferimento ottenuta per il campione x .

- A questo punto, bisogna dividere la Quantity del gene della mesotelina del campione x , per il NF del campione x :

$$Q_{campione x} / NF_{campione x}$$

Questo metodo di analisi statistica viene applicato ai risultati di ciascuno dei 5 esperimenti. I valori di mesotelina normalizzati vengono quindi sottoposti ad analisi statistica con il software dedicato StatGraphics (StatPoint Technologies, Inc.).

In questo protocollo si hanno tre variabili: il trattamento con il microRNA, l'esperimento, ed il tempo di estrazione, e per poterli includere tutti nell'analisi dopo ogni esperimento è stata valutata la variazione dei livelli di espressione del messaggero tra controllo negativo e trattamento nel tempo. In aggiunta, alla fine è stata eseguita una trasformazione (Z-score) sui livelli di mesotelina normalizzata di ogni campione per i cinque esperimenti al fine di eliminare la variabile "esperimento".

La standardizzazione permette di ricondurre una variabile aleatoria distribuita secondo una media μ e varianza σ^2 , ad una variabile aleatoria con distribuzione "standard", di media zero e varianza pari a 1, secondo la formula:

Z-score = $[(X - \mu) / \sigma]$ dove X è il valore di espressione medio del campione x , μ è la media di tutti i valori medi della mesotelina di tutti i campioni (trattamento e controllo) del particolare esperimento, e σ è la deviazione standard tra tutti i valori di mesotelina di tutti i campioni per il medesimo esperimento.

I punti di Z-score di tutti i campioni di tutti gli esperimenti possono essere così comparati in un'analisi statistica ANOVA a due variabili: il livello di espressione della mesotelina (controlli negativi vs trattamenti) e tempo di estrazione.

B.4.2 Analisi dei livelli di espressione della proteina

Per valutare la concentrazione proteica della mesotelina tra i controlli negativi ed i campioni trattati con il microRNA si esegue un Western Blot (WB) su estratti proteici dell'esperimento 2.

Il WB è una tecnica analitica per la visualizzazione di proteine estratte da un campione, che sia un tessuto o un lisato cellulare. Alla base, vi è una elettroforesi su gel di acrilammide (SDS-page) finalizzata alla separazione di proteine denaturate in accordo al loro peso molecolare. Alla migrazione fa seguito il trasferimento delle proteine su membrana, tipicamente di nitrocellulosa, mediante l'applicazione di una corrente elettrica. Il sandwich (che mette a contatto diretto il gel e la membrana) viene immerso in una soluzione tampone attraverso cui viene fatto passare il campo elettrico, così che le proteine, cariche negativamente perché sottoposte al trattamento con SDS, si spostino dal gel alla membrana.

Una volta che le proteine sono state trasferite su membrana, si fa interagire quest'ultima con l'anticorpo specifico per la proteina che si vuole analizzare.

Per la mesotelina si utilizza come anticorpo primario "Mesothelin-G1" (sc271540; Santa Cruz Biotechnology), un anticorpo monoclonale di origine murina diretto contro i residui aminoacidici 295–574, in prossimità del C-terminale della proteina umana.

Come riferimento si utilizza l'actina, la cui espressione stabile fa sì che possa essere utilizzata per la normalizzazione dei livelli di espressione, un po' come se fosse l'housekeeping per le proteine. L'anticorpo utilizzato per la sua visualizzazione è "anti-actin clone C4" (Cat n°MAB1501; Millipore), un anticorpo monoclonale che lega un epitopo situato in una regione altamente conservata della proteina, così da reagire con tutte e 6 le isoforme di vertebrato.

Dal momento che actina e mesotelina hanno un peso molecolare piuttosto simile, bisogna preparare due gel, e quindi due membrane, caricate con i medesimi campioni, una delle quali sarà rilevata con l'anticorpo contro la mesotelina e l'altra con l'anticorpo contro l'actina.

Per la rilevazione vera e propria è infine necessario un anticorpo secondario che interagisca con il primario e che –essendo coniugato con un enzima come una perossidasi o una fosfatasi alcalina, o un fluorocromo– ne permetta la visualizzazione sulla membrana. Entrambi gli anticorpi primari sono stati prelevati da topo, e come anticorpo secondario se ne utilizza uno generico che reagisce specificamente con anticorpi murini, "Goat anti-mouse IgG-HRP" (sc2005; Santa Cruz Biotechnology), coniugato alla perossidasi di rafano.

Quantificazione delle proteine estratte

Nonostante il riferimento fornito dall'actina, per ciascun campione è previsto il caricamento della medesima concentrazione di proteine, che devono quindi essere quantificate.

La quantificazione è stata effettuata allo spettrofotometro, e non avendo indicazioni preliminari sulla resa del protocollo di estrazione seguito, è richiesta la costruzione di una retta di taratura, partendo da campioni a concentrazione proteica nota, così da poter ricavare il range in cui si collocano gli estratti proteici ed eventualmente intensificare i punti per una quantificazione migliore.

Preparazione degli standard per la retta di taratura:

1. 3µl di SDS 1%; replica la condizione del campione, che è infatti disciolto in SDS all'1%.
2. BSA in concentrazioni a scalare; fornisce la concentrazione nota di proteine nello standard.
3. 500µl di working solution (25 parti reagente A, 24 parti reagente B, 1 parte reagente C, in base al volume finale necessario alla quantificazione di campioni, in doppio, e degli standard per la retta di taratura).
4. H₂O sterile a raggiungimento del volume finale = 1000µl.

La BSA è a concentrazione 2mg/ml, e va diluita per portarla alla concentrazione desiderata dopo una diluizione iniziale di 1:20 (BSA 1µg/µl); per la retta di taratura vengono utilizzati 8 punti sperimentali, da una concentrazione iniziale di 0,5µg/µl a 40µg/µl.

Preparazione dei campioni:

- 500µl di working solution;
- 500µl di campione.

Il kit contenente la working solution (QuantumMicro Protein; EuroClone EMP015480) è formulato per fornire un rilevamento di tipo colorimetrico ed un dosaggio di proteine in campioni diluiti aventi una concentrazione di 0,5–20µg/µl. Questo metodo si basa sulla capacità delle proteine di ridurre gli ioni Cu²⁺ a ioni Cu⁺: nel reagente B è contenuto il bicinconinato, che viene trasformato in una sostanza violacea dall'interazione

con gli ioni Cu^+ . In questo modo, l'assorbanza allo spettrofotometro del complesso Cu^+ -(Bicinchoninato)₂ è proporzionale alla concentrazione proteica nel campione diluito. Il reagente A del kit è un buffer alcalino, il reagente B una soluzione contenente acido bicinchoninico, mentre il reagente C è una soluzione contenente solfato di rame.

Dopo l'aggiunta della working solution, gli standard ed i campioni devono essere lasciati per 2h al buio in bagnetto a 37°C, necessari al completo svolgimento delle reazioni; una volta recuperati i campioni è possibile apprezzare la gradazione cromatica sul viola, da più lieve a molto intenso, negli standard a crescente concentrazione proteica.

Lettura allo spettrofotometro

Allo spettrofotometro le proteine vengono lette ad una lunghezza d'onda $\lambda = 562\text{nm}$. Prima si leggono gli standard, in doppietta, quindi si sottrae alla media delle singole doppiette la lettura del bianco, ed i valori ottenuti vengono impiegati per la costruzione della retta di taratura, a formula $y = ax + b$, che ha in ordinate l'assorbanza ed in ascisse la concentrazione proteica. Lo stesso viene fatto per le letture in doppietta dei campioni, e le medie sottratte al bianco vengono utilizzate per ricavare la concentrazione sfruttando l'equazione della retta di taratura:

$$[\text{proteina nel campione } x] = \frac{(\text{valore di assorbanza} - b)/a}{3}$$

Per ogni campione si caricano da protocollo 20µg di proteine in 13µl totali di volume, con l'eventuale differenza tra estratto proteico e volume finale che viene compensata dall'aggiunta di SDS1%.

B.5 Protocollo di Western Blot per 2 gel

B.5.1 Assemblaggio

Montare i supporti per colare il gel e posizzionarli nell'apposito alloggiamento (come raffigurato nell'immagine B.5 a fronte);

Questa fase è molto importante, perché se i vetrini non aderiscono perfettamente, il gel può colare fuori dall'intercapedine inferiore; i gel vengono infatti colati in verticale, e non in orizzontale come i gel di agarosio delle corse elettroforetiche.

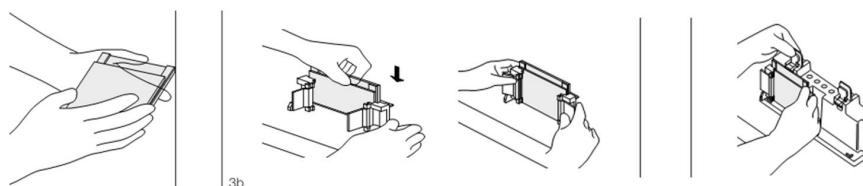


Figura B.5: Assemblaggio delle componenti per il Western Blot: posizionamento dei vetrini per colare i gel; inserzione negli alloggiamenti; posizionamento nel supporto per colare i gel

B.5.2 Preparazione dei gel; in falcon da 15 μ l

Il western blot segue il protocollo di preparazione di gel discontinui di poliacrilamide, ed infatti un singolo gel è composto da due matrici a composizione differente, lo stacking ed il running: il running gel è quello in cui avviene la corsa elettroforetica vera e propria, mentre lo stacking costituisce la matrice in cui vengono inseriti i pettinini, e dove quindi vengono caricati i campioni.

- Composizione dei gel:

GEL	H ₂ O	Gel Buffer	30% Acrilamide/Bis	10%SDS	Vol. totale
Running (12%)	3,4ml	Resolving buffer: 2,5ml	4ml	0,1ml	10ml
Stacking (5%)	2,85ml	Stacking buffer: 1,25ml	0,85ml	0,05ml	5ml

Composizione del Resolving Gel Buffer - 1,5M Tris-HCl, pH8.8 : 1,5M Tris-HCl, pH 8.8 (1L) buffer (Cat n° 161-0798; Bio-Rad).

Composizione dello Stacking Gel Buffer - 0,5M Tris-HCl, pH6.8: 0,5M Tris-HCl, pH 6.8 (1L) buffer (Cat n° 161-0799; Bio-Rad).

Acrilamide/Bis: 30% acrylamide/bis solution, 37,5:1 mixture (30%T, 2,67% C); (Cat n° 161-0156, 500ml; Bio-Rad).

10%SDS: dissolvere 10g di SDS in 90ml di acqua, quindi portare ad un volume finale di 100ml in acqua deionizzata.

- Aggiungere al running gel 50 μ l di APS 10% e 5 μ l di TEMED; l'APS (100mg di persolfato di ammonio dissolto in 1ml di acqua deionizzata) catalizza la polimerizzazione dell'acrilamide assieme al TEMED;
- Agitare il gel e colarlo nell'intercapedine tra i due vetrini di supporto fino a che non raggiunge quasi la sommità; per livellare il gel, aggiungere isopropanolo fino alla sommità dei vetrini;

- Attendere 30min che il gel solidifichi, a Tamb, controllando che il livello del gel non scenda eccessivamente (nel caso, va rifatto da capo);
- Prelevare i vetrini dall'impalcatura di supporto, quindi rovesciare l'isopropanolo e lavare un paio di volte con acqua utilizzando una pipetta pasteur; asciugare con carta assorbente per rimuovere completamente l'acqua rimasta tra i vetrini. Rimettere i vetrini sul supporto;
- Aggiungere allo stacking gel 25 μ l di APS 10% e 5 μ l di TEMED;
- Agitare dolcemente il gel quindi colarlo tra i due vetrini sopra al running gel, fino al bordo superiore, quindi inserire i pettinini che daranno forma ai pozzetti di caricamento;
- Attendere 30min a Tamb che il gel solidifichi;
- Nell'attesa, si preparano i campioni da caricare.

B.5.3 Preparazione dei campioni

- Calcolare il volume in μ l di proteine, corrispondente a 20 μ g, da caricare per ciascun campione;
- Preparare il loading buffer, secondo la proporzione:
50 μ l di β -mercaptoetanololo : 1000 μ l tot = x μ l di β -mercaptoetanololo : (13 μ l * numero di campioni da caricare)
Per campione vengono aggiunti 13 μ l di loading buffer, e la differenza tra il β -mercaptoetanololo ed il volume finale viene compensata dall'aggiunta di SDS reducing buffer.
Composizione del Sample buffer: 3,55ml di acqua deionizzata; 1,25ml di 0,5M Tris-HCl, pH 6.8; 2,5ml di glicerolo; 2,0ml di SDS 10%; 0,2ml di Blu bromofenolo 0,5% (volume totale: 9,5ml).
- Aggiungere 13 μ l di loading buffer a ciascun campione (estratto proteico + SDS 1%);
- Denaturare le proteine tenendo i campioni a 97°C per 10min.

Quello appena descritto è il sistema basato sul laemmli buffer, che viene applicato al protocollo per gel discontinui, in particolare gli SDS-page, dove le proteine vengono denaturate per mezzo del loro riscaldamento in un buffer contenente SDS (sodio dodecil solfato) ed un agente riducente, il β -mercaptoetanololo. In questo modo, le proteine assumono conformazione lineare ed un rapporto carica/massa uniforme e proporzionale al peso molecolare, così che nella corsa su gel vengano separate in accordo ai pesi molecolari.

B.5.4 Caricamento dei campioni

- Una volta che lo stacking gel è solidificato, rimuovere i pettinini e lavare delicatamente i pozzetti con acqua, quindi togliere tutta l'acqua con l'aiuto di carta assorbente;
- Trasferire i vetrini dal supporto di preparazione del gel a quello di caricamento e corsa (come da fig B.6, da 4a a 4d);

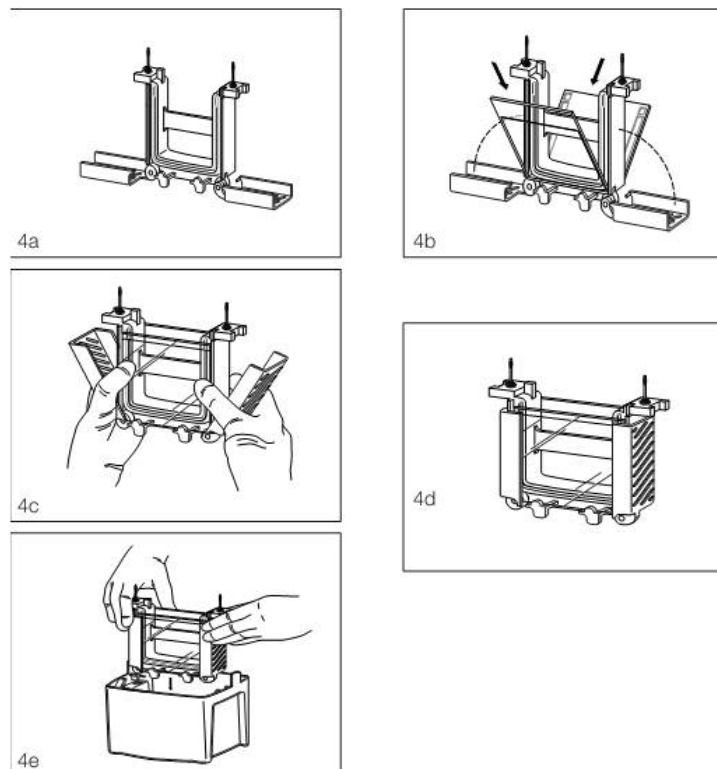


Figura B.6: Alloggiamento dei vetrini nel supporto di corsa: supporto di corsa; posizionamento dei vetrini; chiusura e composizione finale del supporto di corsa

- Inserire l'alloggiamento nella vaschetta di corsa, facendo attenzione ai colori degli elettrodi, mettendo il rosso vicino al rosso, ed il nero vicino al nero (come da fig B.6,4e);
- Aggiungere il buffer di corsa (running buffer: 900ml di acqua deionizzata, e 100ml di 10x Tris/Gly/SDS) prima nell'alloggiamento, controllando che non coli all'esterno (questo porterebbe, durante la corsa, ad un crollo nei livelli di amperaggio, inficiando l'efficienza), quindi riempire tutta la vaschetta fino al livello segnato "2 gel" (550ml circa);

- Caricare nel primo pozzetto 4 μ l di ladder, che come per l'elettroforesi su gel di agarosio dà un riferimento per il peso molecolare delle proteine, quindi 26 μ l totali di campione per pozzetto, facendo attenzione che non ci siano perfusioni in quelli adiacenti;
- Chiudere la vaschetta di corsa, facendo sempre attenzione che le polarità degli elettrodi combacino.

La migrazione delle proteine avviene a voltaggio costante= 120V, per 1h circa. É comunque possibile regolare il tempo di corsa guardando il livello raggiunto dal loading buffer e dal ladder (che ha marcature con coloranti diversi per i diversi pesi molecolari di riferimento, nello spettro del visibile).

L'ampereaggio in questo caso non dovrebbe scendere mai sotto i 15–10V, ma se questo dovesse succedere è indice di un tampone di corsa troppo vecchio, che andrà sostituito con uno nuovo nel successivo western blot.

B.5.5 Trasferimento su membrana

- Spegnere l'emettitore di corrente, quindi aprire la vaschetta di corsa e prelevare l'alloggiamento in cui sono collocati i vetrini; se il tampone di corsa è ancora buono, conservarlo così da poterlo riutilizzare;
- Sciacquare la vaschetta di corsa ed asciugarla;
- Bagnare in una scatoletta le spugnette, le carte e la membrana per il trasferimento con un poco di tampone di trasferimento;
- Prendere il supporto del sandwich, con il lato nero alla base, quindi impilare, nell'ordine:
 - spugna
 - carta grande assorbente
 - carta piccola
 - membrana
 - gel: togliere il gel dai vetrini, quindi rimuovere con una spatolina i dentelli, segno dei pozzetti, rimasti
 - carta piccola
 - carta grande assorbente
 - spugna

NB: per evitare che rimangano delle bolle tra i diversi strati del sandwich, stendere ogni componente aggiunto con un piccolo cilindro bagnato con il tampone di trasferimento.

- Chiudere il sandwich (il lato trasparente sta così in altro) e collocarlo nell'apposito alloggiamento per il trasferimento, quindi mettere il tutto nella vaschetta di corsa con il lato rosso dell'alloggiamento che guarda lo sperimentatore. Il trasferimento avviene in direzione del rosso, quindi i sandwich vanno collocati con il lato nero che guarda il fronte rosso, e quello trasparente rivolto verso il fronte nero;
- Collocare un ghiacciolino nella vaschetta di corsa, quindi aggiungere il buffer di trasferimento (Transfer buffer: 700ml di acqua deionizzata, 200ml di metanolo, 100ml di 10x Tris/Gly Buffer, Bio-Rad, Cat n°161-0734) fino alla tacca "blotting" e chiudere la vaschetta facendo sempre attenzione alle polarità degli elettrodi.

Il trasferimento viene effettuato ad amperaggio costante= 200mA per 1h circa.

B.5.6 Trattamento della membrana

Una volta ultimato il trasferimento (per aver la certezza che tutto si sia svolto correttamente, il ladder deve essersi trasferito dal gel alla membrana), si recuperano i sandwich e si conserva il tampone di trasferimento, che proprio come quello di corsa può essere riutilizzato più volte; le carte grandi e le spugnette vengono tenute, mentre le carte piccole ed i gel possono essere buttati.

Se si conosce il peso molecolare della proteina di interesse, si può tagliare la membrana utilizzando il ladder come riferimento, in modo da ridurre le dimensioni e risparmiare sui volumi dei successivi reagenti.

- Lasciar asciugare all'aria la membrana per 30sec–1min, quindi collocarla in una vaschetta di dimensioni idonee (può essere una piastra petri, o contenitore per puntali) e coprirla con PBS tween;
- Fare due lavaggi di 5min ciascuno in PBS tween sul basculante;
- **Blocking:** incubare per 1h sul basculante le membrane in latte al 5% in PBS tween (per 20ml, 1gram di latte in polvere –Blotting grade blocker non-fat dry milk– disciolto in 20ml di PBS tween).

Questa fase serve per bloccare l'interazione degli anticorpi con eventuali siti aspecifici sulla membrana stessa, che essendo fatta apposta per legare le proteine, ha la capacità di legare anche gli anticorpi. In alternativa al latte può essere utilizzata l'albumina di siero

bovino (BSA), sempre al 5% in PBS tween; entrambe sono soluzioni con proteine diluite, che bloccano i siti liberi sulla membrana, prevenendo così il legame degli anticorpi.

- Fare altri due lavaggi alla membrana di 5min ciascuno sul basculante con PBS tween.

B.5.7 Applicazione dell'anticorpo primario

Come già sottolineato, mesotelina ed actina hanno circa lo stesso peso molecolare (rispettivamente, circa 40kDa e circa 43kDa) quindi sono stati eseguiti due gel, uno per proteina, caricati con i medesimi campioni. L'incubazione prevede l'aggiunta di una quantità precisa di anticorpo primario, che varia a seconda della proteina stessa e della qualità dei campioni, per cui troppo anticorpo può portare a saturazione, mentre troppo poco porta ad un segnale molto debole. Sulla base di indicazioni preliminari, si sono scelte le seguenti diluizioni:

- anti-mesotelina: 1 : 250 in BSA 5% (in PBS tween)
- anti-actina: 1 : 5000 in BSA 5% (in PBS tween)

- Coprire bene le membrane con la BSA al 5%, quindi aggiungere in un angolo, inclinando il contenitore, i μ l di anticorpo corrispondenti alle diluizioni, sulla base del volume finale ottimale alla completa copertura delle membrane (ecco perché più piccola è la membrana e più piccolo è il contenitore in cui può essere alloggiata, e minore sarà il volume totale di soluzione richiesto alla sua copertura). La concentrazione di BSA influisce sulla stringenza del legame dell'anticorpo.
- Lasciare in incubazione sul basculante per 1h a Tamb.
- Tenere le membrane nei corrispettivi recipienti in incubazione overnight a 4°C con l'anticorpo primario

B.5.8 Applicazione dell'anticorpo secondario

Il mattino seguente si recuperano entrambe le soluzioni contenenti l'anticorpo primario, e le si conserva in falcon pulite in frigorifero a 4°C, dal momento che possono essere riutilizzate.

- Fare due lavaggi con PBS tween di 10min ciascuno sul basculante;
- L'anticorpo secondario, "goat anti-mouse", viene diluito secondo un fattore di 1 : 6000 in BSA all'1% (1%BSA + 99% PBS tween). Come per l'anticorpo primario, anche in questo caso viene aggiunto prima un volume idoneo di BSA 1%, e subito dopo i μ l di anticorpo secondario proporzionali al fattore di diluizione;
- Lasciare in incubazione per 1h sul basculante a Tamb;
- Fare due lavaggi in PBS tween di 5min ciascuno sul basculante;
- Fare un lavaggio con PBS (no tween) 1% di 5min sul basculante;
- Aggiungere nuovamente PBS 1% per evitare che la membrana si secchi prima della rilevazione.

B.5.9 Lettura al ChemiDoc MP System (Bio-Rad)

Il sistema presenta una CCD camera per la registrazione delle immagini in tempo reale, così da poter visualizzare il corretto posizionamento della membrana sotto il rilevatore, basato su luce bianca (che permette ad esempio di fotografare il ladder), UV ed altri filtri al LED (per la fluorescenza e la chemiluminescenza), a seconda del tipo di rilevazione che si vuole effettuare.

Il software "Image Lab" permette poi di gestire l'acquisizione e l'analisi dei risultati. Prima della lettura, le membrane devono essere opportunamente trattate per la rilevazione:

- Rimuovere il PBS 1% ed asciugare la membrana all'aria, tamponando gli angoli su carta assorbente;
- Per ogni membrana, preparare in una falcon da 15ml una soluzione composta di 700 μ l buffer Oxidazing reagent + 700 μ l di Enhancer luminol reagent (Kit Western Lightning plus ECL, enhanced chemiluminescence substrate; Perkin Elmer Inc), quindi dispensarla goccia a goccia sulla membrana, dando la priorità all'altezza attesa per la proteina;
- Attendere 1min, lasciando le membrane coperte con carta stagnola, per evitare che la luce infici l'efficienza del luminolo e quindi l'emissione della chemiluminescenza;
- Lasciare asciugare la membrana all'aria, tamponandone eventualmente gli angoli con carta assorbente;
- Procedere alla rilevazione con il ChemiDoc:
 - . Aprire il software e posizionare la membrana sotto la CCD camera;
 - . Selezionare un protocollo di acquisizione single-channel;
 - . Nel setting di acquisizione dell'immagine, selezionare il tipo di colorazione del blot;

- . Definire l'area della membrana da acquisire;
 - . Performare una auto-acquisizione dell'immagine, per bande intense, in modo che sia il programma a stimare il tempo di esposizione ottimale della membrana;
 - . Per la mesotelina, il tempo di auto-rilevazione è di circa 50–60sec, e per l'actina di 20–30sec.
- Sulla base del risultato dell'esposizione automatica, è possibile personalizzare i tempi di esposizione, o scegliere quante foto ottenere per un dato intervallo di tempo, così da poter scegliere la migliore, anche se in questo modo l'esposizione progressiva fa aumentare il rumore di fondo rispetto ad una foto singola scattata con lo stesso tempo di esposizione.
- . L'immagine acquisita può essere ulteriormente modificata, evidenziando ad esempio i punti di saturazione nelle bande, ottenere una visualizzazione 3D delle bande, così che abbiano una altezza proporzionale all'intensità . . .
 - . Salvare ed esportare l'immagine nel formato desiderato.

Descrizione dettagliata delle principali fasi per lo studio evolutivo in silico del gene *MSLN*

C.1 Scelta delle specie

In un primo momento, sono state scaricate le sequenze aminoacidiche e nucleotidiche (DNA e sequenza codificante) della mesotelina umana e dei suoi ortologhi riportati sul database dell'NCBI e dell'Ensembl:

- 10 Primati (*Callithrix jacchus*; *Gorilla gorilla*; *Homo sapiens*; *Macaca mulatta*; *Nomascus leucogenys*; *Otolemur garnettii*; *Pan paniscus*; *Pan troglodytes*; *Papio anubis*; *Pongo pygmaeus*; *Saimiri boliviensis*)
- 7 Roditori (*Cavia porcellus*; *Cricetulus griseus*; *Dipodomys ordii*; *Ictidomys tridecemlineatus*; *Mus musculus*; *Ochotona princeps*; *Rattus norvegicus*)
- 4 Cetacei (*Odobenus rosmarus divergens*; *Orcinus orca*; *Trichechus manatus latirostris*; *Tursiops truncatus*)
- 5 Artiodattili e Perissodattili (*Bos taurus*; *Equus caballus*; *Loxodonta africana*; *Ovis aries*; *Sus scrofa*)
- 3 Carnivora (*Canis lupus familiaris*; *Felis catus*; *Mustela putorius furo*)
- 5 altri Mammiferi (*Ailuropoda melanoleuca*; *Erinaceus europaeus*; *Myotis lucifugus*; *Procavia capensis*; *Pteropus vampyrus*)
- 2 Marsupiali (*Monodelphis domestica*; *Sarcophilus harrisii*)
- 2 Uccelli (*Gallus gallus*; *Meleagris gallopavo*)
- 2 Rettili (*Anolis carolinensis*; *Pelodiscus sinensis*)
- 1 Anfibia (*Xenopus tropicalis*)

- 3 Pesci (*Danio rerio*; *Latimeria chalumnae*; *Oryzias latipes*)

L'elenco di partenza ha dovuto subire una decisa scrematura per rimuovere le sequenze incomplete o con posizioni non definite (ad uno o più dei tre livelli di informazione genica) e quelle presentanti “frameshift exons”, coppie di esoni ed introni dalla lunghezza inferiore alle 20bp e con siti accettori e donatori di splicing non canonici, generati da predizioni bioinformatiche per mantenere il quadro di lettura della proteina.

Un secondo livello di selezione è quello che ha portato all'elenco definitivo di specie mostrato nell'introduzione (“La mesotelina umana ed i suoi orologi”, tab. 1.1 a pagina 15): sono state mantenute unicamente le sequenze di specie appartenenti alla classe Mammalia, dal momento che quelle degli altri taxa mostrano una divergenza evolutiva tale da andare a coprire il segnale evolutivo di interesse per l'analisi.

C.2 Generazione degli allineamenti multipli ed analisi delle sequenze

Definizione dei confini esone-esone

Al di là della conservazione della sequenza, un dato molto importante da analizzare è se anche la fase e la lunghezza dei singoli esoni è mantenuta a livello comparato; a tal fine, la sequenza completa del cDNA è stata sottoposta al software “Translate tool” disponibile sul sito dell'ExpASy (<http://web.expasy.org/translate/>)

- Per ciascuna specie, la sequenza codificante del messaggero è stata tradotta in silico, inizialmente impostando il filtro “Verbose; Met, Stop”, che riporta le diverse traduzioni possibili a seconda che la lettura sia fatta dal 5' al 3' (senso) o dal 3' al 5' (antisenso), ed in base alla fase (0, 1 o 2) della prima tripletta. Una volta identificato il quadro di lettura corretto, si ripete la traduzione in silico selezionando come filtro “Include nucleotide sequence”. In questo modo, per ogni tripletta di nucleotidi della sequenza viene restituito il codone e la traduzione corrispondente, che vanno copiati ed incollati in un file word dedicato alla specie.
- La definizione dei siti di giunzione esone-esone viene effettuata manualmente, utilizzando come banca dati genomica di riferimento l'UCSC genome browser, e nello specifico il BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>): è sufficiente inserire

la sequenza del cDNA e selezionare dall'elenco la specie con la quale fare il confronto, controllare che la release sia quella più aggiornata, e quindi sottomettere la ricerca.

Il database ricerca, tra tutte quelle depositate per la specie selezionate, le sequenze corrispondenti a quella sottomessa, indicando anche la percentuale di omologia e la lunghezza, in nucleotidi, del dato riportato. Tra i risultati, si sceglie quello di interesse, e si controlla che la sequenza sottomessa combaci con quella depositata (eventuali differenze rispetto a quella depositata sono segnate con lettere in minuscolo ed in nero). Il dato più importante è quello della sequenza genomica, dove gli esoni sono segnati in maiuscolo ed in blu e gli introni in minuscolo e nero; in più, viene restituita la corrispondenza tra triplette ed amminoacido ("side by side alignment").

- Sfruttando la distinzione tra esoni ed introni dell'assemblaggio genomico, bisogna cercare, nella traduzione in silico, i confini di ogni esone, evidenziando nella tripletta quali nucleotidi appartengono ad un esone e quali a quello successivo, a seconda della fase:
 - Fase 0) Gli ultimi tre nucleotidi dell'esone rappresentano una tripletta, ossia un codone, definente l'ultimo amminoacido dell'esone stesso;
 - Fase 1) L'ultimo nucleotide dell'esone corrisponde al primo di una tripletta, che sarà quindi completata dai primi due nucleotidi dell'esone successivo; in questo caso, l'amminoacido viene considerato come appartenente al secondo esone, per quanto un nucleotide della tripletta sia dell'esone precedente;
 - Fase 2) Gli ultimi due nucleotidi dell'esone rappresentano i primi due di una tripletta, che è completata dal primo nucleotide dell'esone successivo; l'amminoacido in questione viene considerato come appartenente al primo esone.
- Nel file con l'allineamento proteico, indicare con una barra "I" la divisione esone per esone per ogni specie, la numerazione degli esoni, e la fase di fine di ciascuno; confrontare se questi corrispondono per tutte le specie, o se per alcuni vi sono divergenze.
- Mentre si segnano le fasi, è importante controllare anche se vengono rispettati i siti accettori e donatori di splicing canonici: gli introni devono iniziare con un doppietto GT, e terminare con un doppietto AG/GC.
- Lo stesso protocollo viene quindi ripetuto per gli allineamenti del cDNA e della sequenza genomica, anche se la maggiore divergenza, in particolare dovuta alla differente lunghezza degli introni tra specie e specie, ha reso necessarie più correzioni manuali negli allineamenti.

La comparazione della struttura esoni-introni ha rappresentato uno strumento fondamentale per la correzione di alcuni errori di assemblaggio delle sequenze riportate sui database dell'NCBI e dell'Ensembl.

Già dalla letteratura (Kim & Hahn, 2011), ma ancor più dall'analisi diretta delle sequenze del gene della mesotelina, è emerso come all'interno dei Mammiferi sia possibile riconoscere due cluster, o blocchi, principali. Nei primati e nei roditori (*Euarchontoglires*) il gene ha 17 o 16 esoni, in base alla lunghezza della 5'UTR e della 3'UTR, di cui 16 sono quelli codificanti, ma negli altri Mammiferi, formalmente i *Laurasiatheria*, gli esoni codificanti sono 10, a causa di una delezione segmentale che ha rimosso dal gene ancestrale gli esoni 3-8.

Delezioni di tale portata rendono le sequenze ingestibili per qualunque algoritmo, che sia per la generazione di allineamenti, alberi, o altre analisi evolutivo-comparate, e per questo motivo in un primo momento si è deciso di lavorare con due blocchi distinti, il primo relativo agli *Euarchontoglires* (*Cricetulus griseus*; *Homo sapiens*; *Macaca mulatta*; *Mus musculus*; *Nomascus leucogenys*; *Pan troglodytes*; *Papio anubis*; *Rattus norvegicus*; *Saimiri boliviensis*), ed il secondo relativo ai Laurasiateri a confronto con *Homo sapiens* (*Bos taurus*; *Canis lupus familiaris*; *Equus caballus*; *Felis catus*; *Mustela putorius furo*; *Odobenus rosmarus divergens*; *Sus scrofa*).

C.2.1 Generazione degli alberi filogenetici

Gli alberi filogenetici sono stati generati utilizzando il programma "Mega5" (<http://www.megasoftware.net/>).

Sia che si esegua l'analisi di una sequenza aminoacidica, sia che si tratti di una sequenza nucleotidica, la modalità è simile, cambiando solamente gli algoritmi, e quindi le impostazioni interne al programma:

Generazione del file di input .meg

Per ogni specie si deve ricavare dall'allineamento multiplo una sequenza continua, in cui l'allineamento relativo è sempre rispettato, rimuovendo tutti i segni di interruzione "I" tra esone ed esone.

Per importare su Mega5 gli allineamenti bisogna selezionare "Align > Edit/Built Alignment > create a new alignment" specificando che si vuole analizzare una sequenza

proteica (“protein”).

Uno dei vantaggi offerti dalla veste grafica del programma è la distinzione dei diversi amminoacidi sulla base della classe di appartenenza, così che anche ad una prima occhiata si può ricavare se vi sono blocchi a proprietà chimico-fisiche analoghe.

Analisi della sequenza proteica

Si seleziona il menù “Phylogeny”, e si sceglie il tipo di metodo che si desidera utilizzare per la costruzione dell’albero; sono disponibili metodi che si basano tanto su matrici delle distanze (UPGMA, minima evoluzione e Neighbor joining) quanto su matrici di caratteri (Maximum parsimony e Maximum likelihood). I primi due sono oggi poco utilizzati, fondandosi su algoritmi che pongono troppe assunzioni di base e che portano quindi a risultati poco affidabili, per quanto poi risultino statisticamente robusti. Il neighbor joining è ancora utilizzato, in particolare in fasi preliminari di studio, ma è il metodo di massima verosimiglianza quello al momento reputato più affidabile, e che è stato quindi scelto per l’analisi.

La robustezza degli alberi viene testata per mezzo del bootstrapping (500 repliche), un test statistico per testare i risultati dell’indagine filogenetica che consiste nella creazione di nuovi set di dati (mantenendo però l’albero di partenza) campionando quello originario in modo casuale e con repliche. Se ogni colonna nell’allineamento multiplo corrisponde ad un carattere, il set risultante avrà la stessa dimensione di quello di partenza, ma i caratteri saranno stati disposti secondo un ordine diverso, esclusi e/o replicati, dal momento che il rimescolamento avviene in maniera casuale e con reinserimento del carattere appena estratto.

L’algoritmo elabora un certo numero di allineamenti multipli simulati che vengono analizzati con gli stessi metodi di quello reale, e per ciascuno viene quindi costruito un albero filogenetico. Per la ricostruzione dell’albero consenso l’algoritmo calcola per ciascun nodo la percentuale di alberi simulati in cui viene riprodotto lo stesso nodo (due alberi riproducono lo stesso nodo quando esso condivide le stesse unità tassonomiche discendenti). La percentuale così calcolata viene definita come il valore di bootstrap del nodo considerato. Un alto valore di bootstrap corrisponde ad un’alta significatività statistica del nodo, e di norma possono essere considerati robusti nodi con una percentuale almeno del 50%, se si applica la “majority rule consensus”.

La tecnica del bootstrap dà informazioni sulla precisione, ma non sull’accuratezza del-

l'analisi, fornendo una stima della topologia dell'albero: se il processo evolutivo si è manifestato in modo uniforme per tutta l'estensione della sequenza considerata, i valori di bootstrap tenderanno ad essere più alti. Al contrario, se parti della sequenza hanno subito processi evolutivi differenti, i valori di bootstrap tenderanno ad essere molto bassi.

La seconda impostazione da specificare è il modello da far applicare; si è scelto il WAG* (Whelan & Goldman, 2001) con stima delle frequenze, che è al momento uno tra i più complessi per la costruzione della matrice di sostituzione aminoacidica. Si implementa poi l'analisi con la funzione gamma (Γ), che ha come caratteristica l'ammette eterogeneità nei tassi di sostituzione tra siti, che vengono così ripartiti in 5 categorie principali discrete.

Per gli altri parametri, si lasciano quelli di default:

```
Analysis
  Analysis ----- Phylogeny Reconstruction
  Statistical Method ----- Maximum Likelihood
Phylogeny Test
  Test of Phylogeny ----- Bootstrap method
  No. of Bootstrap Replications -- 500
Substitution Model
  Substitutions Type ----- Amino acid
  Model/Method ----- WAG with Freqs. (+F) model
Rates and Patterns
  Rates among Sites ----- Gamma Distributed (G)
  No of Discrete Gamma Categories - 5
Data Subset to Use
  Gaps/Missing Data Treatment ---- Use all sites
Tree Inference Options
  ML Heuristic Method ----- Subtree-Pruning-Regrafting
Extensive (SPR level 5)
  Initial Tree for ML ----- Make initial tree automatically
(Default - NJ/BioNJ)
  Branch Swap Filter ----- Very Strong
System Resource Usage
```

Il programma genera un albero filogenetico che non assume un modello di evoluzione molecolare (molecular clock), e che risulta quindi senza radice; l'ancestore –ossia la radice– può essere inserita in un punto qualunque sui rami dell'albero. Le lunghezze dei rami rappresentano invece vere distanze evolutive tra specie (il segmento rappresenta l'unità di misura per il numero di sostituzioni rilevate per il ramo specifico), ed accanto a ciascun nodo sono riportati i valori di bootstrap calcolati.

Analisi della sequenza di cDNA

Per analizzare una sequenza nucleotidica i passaggi sono i medesimi elencati per la generazione di file di input per la sequenza proteica, solo che al posto di specificare “Protein” si seleziona “DNA”, e si immettono le sequenze di cDNA ricavate dagli allineamenti.

Una volta creati i file di input .meg è sufficiente selezionare “Create/Test Maximum Likelihood Tree” nel menù Phylogeny, e specificare il modello di analisi; come per le proteine, viene eseguita una analisi di maximum likelihood con 500 repliche di bootstrap secondo i seguenti parametri:

```
Analysis
  Analysis ----- Phylogeny Reconstruction
  Statistical Method ----- Maximum Likelihood
Phylogeny Test
  Test of Phylogeny ----- Bootstrap method
  No. of Bootstrap Replications -- 500
Substitution Model
  Substitutions Type ----- Nucleotide
  Model/Method ----- General Time Reversible model
Rates and Patterns
  Rates among Sites ----- Gamma Distributed (G)
  No of Discrete Gamma Categories - 5
Data Subset to Use
  Gaps/Missing Data Treatment ---- Use all sites
Tree Inference Options
  ML Heuristic Method ----- Subtree-Pruning-Regrafting
Extensive (SPR level 5)
  Initial Tree for ML ----- Make initial tree automatically
(Default - NJ/BioNJ)
  Branch Swap Filter ----- Very Strong
System Resource Usage
  Codons Includedm ----- 1st+2nd+3rd+Non-Coding
```

Il GTR (General Time Reversible) è il modello computazionale più avanzato (Tavaré, 1986), poiché include come casi speciali tutti i modelli precedentemente proposti (JC69, K80, F81, HKY85..) ed ammette la possibilità che i tassi di mutazione siano diversi per ogni coppia di nucleotidi, e che le frequenze all’equilibrio siano differenti per tutte e quattro le basi.

Anche in questo caso, si ottengono alberi senza radice, in cui le lunghezze dei rami corrispondono alla distanza evolutiva tra le specie.

Introduzione dell'outgroup

Per ampliare l'analisi e renderla più solida si è scelto di aggiungere una specie che servisse da outgroup.

Per definizione, in cladistica, con "outgroup" si intende un gruppo monofiletico di organismi (o una singola specie) che non appartiene a quelli in analisi, ma funge da riferimento nella determinazione delle relazioni evolutive tra altri gruppi monofiletici. Nel caso dell'analisi di sequenze, per essere reputata valida, la sequenza deve essere sufficientemente imparentata con quelle in analisi affinché possa esservi comparata, ma meno strettamente di quanto non lo siano le stesse tra loro.

Tra le specie non appartenenti alla classe Mammalia, nessuna ha una sequenza per la mesotelina sufficientemente simile (per quanto riguarda gli uccelli il problema è legato non solo alla datazione della divergenza, ma alle peculiarità evolutive delle sequenze, che mostrano spesso tassi di mutazione poco concordi con il tempo di separazione evolutivo). Per questo motivo, si è provato a guardare tra marsupiali e monotremi (la cui divergenza è meglio ricondotta al Mesozoico, tra Triassico e Giurassico) per quali specie fosse disponibile nei database una sequenza sufficientemente definita per la mesotelina. L'unica specie selezionabile è l'opossum (*Monodelphis domestica*), per quanto l'omologia di sequenza è molto limitata, in particolare per i primi esoni ed introni (il gene ha 28 esoni, molti dei quali assenti nei placentati).

E' stato quindi ripetuto l'allineamento a blocchi includendo in entrambi la sequenza dell'opossum, analizzata in relazione a quella umana, quindi sono stati generati gli alberi filogenetici.

Generazione degli alberi filogenetici con Phylip

Sia che si esegua l'analisi di una sequenza aminoacidica, sia che si tratti di una sequenza nucleotidica, la modalità è simile, cambiando solamente gli algoritmi, e quindi gli eseguibili:

Generazione del file di input "infile"

Si partono dai documenti degli allineamenti in formato .txt già sfruttati per l'analisi con Mega5, in cui erano stati preventivamente rimossi tutti i segni di interruzione "I" tra esone ed esone.

La prima riga del file deve contenere in questo caso il numero di specie e la lunghezza della sequenza, che derivando da un allineamento, deve essere la stessa per tutte le specie. Ad esempio, per la proteina del blocco 1, nella prima riga sarà indicato: “ 9 676”

Analisi della sequenza proteica

1. Lanciare l'eseguibile “protdist.exe”, che calcola una matrice di distanze evolutive tra sequenze secondo cinque possibili modelli di sostituzione aminoacidica, tenendo di conto di possibili gap “-” o ambiguità “?” . La distanza per ciascuna coppia di specie stima la lunghezza totale del braccio tra le stesse. Come modello è stato scelto il Categories distance, che è stato formulato dai programmatori del programma per sopperire ad alcune limitazioni intrinseche di altri modelli; l'idea di base è quella di una sequenza nucleotidica che varia in accordo al modello di Kimura a due parametri (con differente tasso di transizioni contro le trasversioni), con la differenza tuttavia che alcune variazioni aminoacidiche sono meno probabili di altre, motivo per cui gli amminoacidi vengono raggruppati in una serie di categorie distinte, come la classificazione di George-Hunt-Barker (1988), quella di Ben Hall, o quella di Conn e Stumpf (1963).
2. Convertire l'outfile in un nuovo infile.
3. Lanciare l'eseguibile “neighbor.exe”, che costruisce un albero attraverso un processo di clusterizzazione progressiva delle linee filetiche definendo al tempo stesso la lunghezza dei rami e le diramazioni. Analogamente a Mega5, l'albero non assume un modello di evoluzione molecolare (molecular clock), motivo per cui quello che viene generato è un albero senza radice. Come modelli vengono forniti sia l'UPGMA che il Neighbor joining, ma si consiglia di utilizzare come modello il secondo, che partendo dalla matrice di distanze fornita nell'infile (in cui sono specificate le distanze tra ciascuna coppia di taxa) e da un albero filogenetico non risolto dalla topologia a stella, in cui tutti i rapporti sono equiprobabili, itera un algoritmo secondo cui si trova ad essere associata quella coppia di taxa a distanza minore. Aggiungendo di volta in volta un nuovo gruppo al cluster già risolto, viene ricostruito l'albero che prevede il totale di lunghezza dei rami più basso, e dove la lunghezza dei rami è proporzionale agli eventi evolutivi che separano le singole specie.
4. Aprire l'outree con il programma “treeview” e salvarlo in formato grafico; salvare l'outree come file .txt, perché servirà come file di input per le analisi successive.

Non è necessario ripetere l'operazione per generare alberi con valori di bootstrap, poichè servono unicamente la topologia e la lunghezza dei rami.

Analisi della sequenza di cDNA

1. Lanciare l'eseguibile "dnadist.exe"; il programma utilizza sequenze nucleotidiche per calcolare una matrice di distanza in accordo a quattro possibili modelli di sostituzione nucleotidica in cui la distanza per ciascuna coppia di specie stima la lunghezza totale del ramo tra le due. E' stato scelto come modello di sostituzione quello proposto da Kimura nel 1980, a due parametri, circa simmetrico nelle probabilità di sostituzione tra nucleotidi, in cui tuttavia è contemplato due tassi diversi per transizioni e trasversioni:

To:	A	G	C	T

A	1-a-2b	a	b	b
From: G	a	1-a-2b	b	b
C	b	b	1-a-2b	a
T	b	b	a	1-a-2b

“a” è il prodotto del tasso di sostituzione per unità di tempo e della lunghezza dell'intervallo di tempo trascorso (valido per tempi di divergenza “brevi”), mentre “b” è il prodotto tra metà del tasso di trasversione e della lunghezza dell'intervallo di tempo. Ciascuna distanza calcolata rappresenta una stima, per quella particolare coppia di specie, del tempo di divergenza tra queste; l'algoritmo applicato differisce tuttavia leggermente da quello originale di Kimura, poichè permette una computazione della stima di maximum likelihood dipendente dalle sequenze osservate solo sulla base della loro lunghezza e del numero osservato di differenze tra transizioni e trasversioni tra coppie di sequenze.

2. Convertire l'outfile in infile.
3. Ripetere lo stesso protocollo di analisi utilizzato per la sequenza proteica.

C.3 Analisi filogenetica ed evolutiva

Al fine di testare se in particolari linee evolutive la mesotelina è andata incontro a processi di selezione è stato impiegato PAML –Phylogenetic Analysis by Maximum Likelihood– (<http://abacus.gene.ucl.ac.uk/software/paml.html>; Yang), un pacchetto di programmi strutturato sul metodo di maximum likelihood. Questo ha la grande potenzialità

di utilizzare direttamente le sequenze che vengono sottomesse per la generazione delle matrici di sostituzione, ed è quindi ancora più sensibile ad allineamenti con bassi livelli di omologia e gap. Il metodo cerca di quantificare la probabilità che ad una ipotesi H (l'albero filogenetico) corrisponda un certo set di dati D (l'allineamento multiplo). La probabilità viene calcolata come il prodotto delle probabilità che ha ciascun sito dell'allineamento multiplo di presentare un certo carattere se su questo ha agito un processo evolutivo. Tutte le variabili del modello vengono computate per massimizzare il valore di verosimiglianza (likelihood) della ricostruzione filogenetica.

E' stato applicato il programma "codeml", impostando come dati per l'analisi i codoni della sequenza di cDNA. Viene effettuata una comparazione tra modelli sito-specifici, ciascuno basato su differenti assunzioni per quanto riguarda il valore ω ($=dN/dS$), indice utilizzato come parametro per la determinazione di pressione selettiva tra siti.

C.3.1 Preparazione dei file di input

Per l'analisi sono necessari tre file, due di input ed uno di istruzioni per impostare i parametri dell'analisi:

- seqfile: rappresenta il file di input in cui è contenuto l'allineamento multiplo. A patto che siano sotto l'ordine della decina, i gap possono essere mantenuti, ma vengono trattati come siti di incertezza (indicati come "?"). Se sono troppi, può essere necessario dire al programma di non considerare i siti non definiti, oppure li si può eliminare direttamente nel file di input, perdendo in entrambi i casi l'informazione relativa. Essendo l'analisi condotta su codoni, la sequenza dovrà essere necessariamente lunga un numero multiplo di 3, e deve essere infine rimosso il codone di stop alla fine della sequenza nucleotidica.
- treefile: è il file di input relativo all'albero filogenetico. Questo deve essere generato da un programma apposito, Phylip.
- codeml.ctl: è questo il file da compilare per specificare le modalità di analisi filogenetica.

Di base, il documento ha il seguente aspetto:

```
seqfile = nome dell'infile * sequence data file name
treefile = nome del treefile * tree structure file name
outfile = nome dell'output * main result file name
noisy = 9 * 0,1,2,3,9: how much rubbish on the screen
verbose = 1 * 1: detailed output, 0: concise output
runmode = 0 * 0: user tree; 1: semi-automatic; 2: automatic;
          3: StepwiseAddition; 4,5):PerturbationNNI;
```

```

-2: pairwise
seqtype = 1 * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 1 * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table
clock = 0 * 0: no clock, unrooted tree, 1: clock, rooted tree
aaDist = 0 * 0:equal, +:geometric; -:linear
model = * models for codons:
* 0:one, 1:b, 2:2 or more $dN/dS$ ratios for branches
NSSites = * 0:one w; 1:NearlyNeutral; 2:PositiveSelection;
3:discrete; 4:freqs; 5:gamma; 6:2gamma; 7:beta;
8:beta&w; 9:beta&gamma; 10:3normal
icode = * 0:standard genetic code; 1:mammalian mt; 2--10:see below
fix_kappa = 0 * 1: kappa fixed, 0: kappa to be estimated
kappa = 1,6 * initial or fixed kappa
fix_omega = * 1: omega or omega_1 fixed, 0: estimate
omega = 0,6 * initial or fixed omega, for codons or codon-based AAs
fix_alpha = 0
alpha = 0
Malpha = 0
getSE = 1 * 0: don't want them, 1: want S.E.s of estimates
RateAncestor = 0 * (0,1,2): rates (alpha > 0) or ancestral states (1 or 2)
Small_Diff = .45e-6
cleandata = 0 * remove sites with ambiguity data (1:yes, 0:no)?
method = 0 * 0: simultaneous; 1: one branch at a time
fix_blength = 0 * 0: ignore, -1: random, 1: initial, 2: fixed

```

Con “model” si specifica il modello di analisi, e con “NSsite” le variabili del modello, che sono i parametri determinanti assieme ad “omega”, che può essere imposto come costante per tutti i rami o stimato dal programma.

Vista la notevole divergenza determinata dall’inserimento dell’outgroup nei primi esoni, e la delezione degli esoni 3–8 nei laurasiateri, si è deciso di raggruppare gli allineamenti dei due blocchi in un unico file includente le 17 specie selezionate, includendo tuttavia solo gli esoni dal 9 al 16 (10–17, se si considera che l’esone 1 di uomo è stato rimosso poichè 5’UTR), che contengono la porzione di precursore proteico codificante per la mesotelina matura. La scelta è stata giustificata dall’analisi della presenza di siti di taglio per enzimi furina-simili utilizzando il programma ProP 1.0 Server (<http://www.cbs.dtu.dk/services/ProP/>), ritrovando un certo consenso una regione assimilabile all’esone 9 (aa 295–297 per primati e roditori e 73–75 per i laurasiateri).

Il “site model” permette di valutare se uno o più codoni sono stati sottoposti a selezione nel corso dell’evoluzione; per rendere l’analisi più sensibile al segnale evolutivo, si è pensato di testare gli esoni a gruppi, e poi singolarmente. Dal momento la lunghezza della sequenza di input deve essere un multiplo di tre, e che l’E9 inizia con una fase

0, per trarre una prima indicazione sulla presenza di siti sottoposti a selezione si sono analizzati blocchi di esoni che rispettassero un inizio ed una fine di fase 0.

In particolare: E9–10 (279nt); E11 (156nt); E12–13–14 (366nt); E15–16 (273nt).

Per l'analisi esone per esone si è aggiustata la suddivisione in accordo a quella precedentemente ricostruita per l'allineamento proteico, spostando da un esone all'altro alcuni nucleotidi per avere sequenze multiple di 3: per gli esoni che terminano con una fase 1 l'ultimo nucleotide è stato spostato all'esone successivo; per quelli di fase 2 invece il primo nucleotide dell'esone successivo è stato spostato a quello di interesse.

C.3.2 Analisi Site Model

Sono stati eseguiti cinque modelli evolutivi in una singola computazione:

- M0 (Model 0, NSsite); è un modello base in cui il valore ω è fisso per tutti i rami, ma viene trattato come parametro libero, l'unico nella distribuzione, stimato sulla base delle sequenze fornite.

Il tasso di sostituzione dal codone i al codone j viene specificato in accordo al tipo di sostituzione evidenziata (transizioni vs transizioni, e per ciascuna se sinonima o non sinonima), mentre la frequenza di equilibrio del codone j (π_j) è stata calcolata dalla media delle frequenze nucleotidiche.

- M1a (Model 0, NSsite 1); è un modello di quasi-neutralità in cui ω può assumere valori $0 < \omega < 1$ secondo quanto stimato sulla base delle sequenze.
- M2a (Model 0, NSsite 2); è questo un modello di selezione positiva in cui vengono accettati tre parametri ω , $\omega_0 < 1$, $\omega_1 = 1$ ed $\omega_2 > 1$, a cui corrispondono tre proporzioni di classi di siti.
- M7 (Model 0, NSsite 7); è un modello di distribuzione beta in cui non sono ammessi siti sottoposti a selezione positiva.
- M8 (Model 0, NSsite 8); è sempre un modello di distribuzione beta, in cui sono ammessi siti sottoposti a selezione positiva.

In maniera molto schematica, si può assumere che un $\omega < 1$ indichi la presenza di selezione purificante (negativa) che previene la modificazione del codone perché sottoposto a vincolo funzionale; un $\omega = 1$ indica evoluzione neutrale, ed $\omega > 1$ la presenza di selezione positiva per il cambiamento aminoacidico o nucleotidico.

Selezionando tutti e 5 i modelli in un singolo run, il programma esegue in automatico

due test di rapporto di verosimiglianza (LRT likelihood ratio tests, approfonditi nell'appendice H a pagina 183) per la presenza di selezione positiva in cui un modello nullo –dove $\omega \leq 1$ – viene comparato ad un modello alternativo in cui omega è libero di assumere qualunque valore. Il primo test compara il modello M1a di quasi neutralità con il modello M2a di selezione positiva, ed il secondo compara il modello M7 con M8, con il primo test che viene indicato come più robusto (o meno potente).

Se emerge un qualche dato indicativo al LRT, viene inferita la presenza di pressione selettiva positiva.

A questo risultato viene accompagnato un test per calcolare la probabilità a posteriori che ciascun sito provenga da una particolare classe di siti secondo i modelli NEB (naive empirical Bayes) e BEB (Bayes empirical Bayes). Tra i due, è più affidabile il BEB, che assegna una probabilità a priori ai parametri del modello, e la integra per le loro incertezze, restituendo la posizione e l'identità dell'amminoacido sottoposto a selezione, la media approssimata della distribuzione a posteriori per ω ed il relativo errore standard. In questo modo, è stato possibile inferire se in alcuni esoni sono presenti siti sottoposti a selezione positiva.

Sempre in accordo con l'analisi site model, si è eseguito il modello 3 (M3 discrete; model=0, NSsite=3), in cui si ammette una pressione selettiva variabile tra siti a 3 parametri liberi nella distribuzione omega: ω_0 , ω_1 , ω_2 . Si esegue quindi il LRT tra i modelli M0, che rappresenta l'ipotesi nulla, ed il modello M3 quale test di variabilità di ω tra siti, piuttosto che come test per la presenza di selezione positiva.

È stato infine eseguito un quarto test per comparare l'ipotesi nulla del modello M8a (che segue sempre una distribuzione beta, ma ha un valore di ω fisso =1, specificato con NSsite=8, fix_omega=1, omega=1) con il modello M8 (NSsite=8, fix_omega=0, omega=0,6).

C.3.3 Analisi del significato delle sostituzioni aminoacidiche

Per capire se le variazioni significative individuate hanno portato a conseguenze nell'amminoacido codificato che influiscono sulla struttura e/o funzionalità della proteina, si è utilizzato il programma SIFT (<http://blocks.fhcrc.org/sift/SIFT.html>). Questo sfrutta un algoritmo basato sull'omologia di sequenza che ricerca sul database di BLAST le

sequenze omologhe a quella sottomessa; oltre alla sequenza di interesse, bisogna specificare le sostituzioni che si vuole valutare. Il programma restituisce un score che dipende dagli amminoacidi che compaiono nell'allineamento multiplo e dalle loro proprietà chimico-fisiche. Lo score rappresenta la probabilità normalizzata che il cambiamento amminoacidico sia tollerato, con un valore soglia di 0,05 al di sotto del quale la sostituzione viene predetta come deleterea per la funzione o la struttura della proteina.

Nel caso in cui una sostituzione viene predetta come deleterea, la confidenza dipende dalla divergenza tra le sequenze che compaiono nell'allineamento: se hanno un alto livello di omologia, allora molte posizioni possono risultare conservate, e come conseguenza anche il numero di sostituzioni non tollerate sarà più alto, con molti falsi positivi (sostituzioni in realtà neutrali, ma predette come deleteree).

Per questo motivo, viene anche calcolato il valore di conservazione media, una misura della diversità delle sequenze nell'allineamento. La conservazione è computata per ciascuna posizione, e da qui viene ricavato il valore medio: quando un sito è completamente conservato, il suo valore è di 4,32 ($\log_{20} 20$), mentre è pari a 0 quando si possono riscontrare tutti e 20 gli amminoacidi nella medesima posizione.

Il possibile stato ancestrale della sequenza è stato ricavato con il tool "Ancestors > Infer Ancestral Sequences" (Maximum Likelihood) del programma Mega5. È sufficiente fornire un file di allineamento multiplo, e quindi tra i parametri scegliere il relativo albero precedentemente generato, il tipo di sostituzione, il modello da applicare (ancora il WAG con frequenze), ed una distribuzione gamma a 5 categorie discrete.

L'eseguibile restituisce l'albero filogenetico, ed indica per ogni nodo l'amminoacido presente nello stato ancestrale per ciascuna posizione, ammettendo incertezze quando non è chiaramente identificabile; per questo motivo, nel sottomettere l'amminoacido di partenza e quello di sostituzione all'analisi con SIFT si è ricorsi allo stato presente nei primati (che quasi sempre presentano uno stato ancestrale dell'amminoacido certo, e clusterizzano indipendentemente dagli altri taxa, come sister group) come prima indicazione, ed alle altre varianti rilevate come sostituzioni. Per maggiori conferme è stato poi fatto l'opposto.

Le indicazioni ricavate da SIFT sono state confrontate (e confermate) con i risultati della stessa analisi condotta con altri tools bioinformatici, come PANTHER PSEC, che sfrutta sempre le informazioni ricavate dalla comparazione di sequenze omologhe, e

con SNPs3D, che sfrutta 15 fattori strutturali e la conservazione di sequenza per fare la predizione.

C.3.4 Stima del valore Omega per il gene *MSLN*

Per avere indicazioni del rate evolutivo del gene *MSLN* si sono analizzati anche i valori di distanza evolutiva a coppie di sequenze per l'intero cDNA e per la porzione di sequenza compresa tra gli esoni 9 e 16.

Per la sequenza ridotta, si è fatto riferimento ai valori di ω restituiti dal Site-Model (PAML), nello specifico:

- Valori medi per tutte le specie sottomesse in accordo ai modelli nullo (M0), M1, M2, M7 ed M8;
- Valori a coppie di specie in accordo al modello formulato da Nei & Gojobori (1986).

Lo stesso tipo di analisi è stato condotto anche con Mega5 per la sequenza completa del cDNA, seguendo il protocollo di Mega5:

- “distance > compute pairwise distance”.
- “Syn-Nonsynonymous” in accordo al modello Nei-Gojobori per uniformità rispetto all'analisi condotta con PAML.
- Calcolare prima le sostituzioni non sinonime (dN), quindi le sinonime (dS).
- Calcolare il rapporto dN/dS per tutte le coppie di sequenze.

Lo scopo di questa analisi in silico è fornire maggiori informazioni sulla mesotelina, tanto a livello di gene quanto di proteina, sfruttando le potenzialità dell'analisi filogenetica comparata per cercare di capire come è evoluta nel tempo la sequenza codificante, se la filogenia rispetta la filogenesi, o ancora se è andata incontro a cambiamenti importanti in alcune linee filetiche.



Risultati dei principali passaggi dello studio funzionale in vitro dell'interazione tra il microRNA miR-204-5p e la mesotelina

Sono di seguito riportati i risultati dei principali passaggi ed esperimenti condotti al fine di testare l'efficacia del microRNA miR-204-5p nel legare la mesotelina e quindi regolarne l'espressione in cellule della linea Mero-14, dove la proteina è molto espressa. Sono stati analizzati tanto i livelli di messaggero quanto di proteina presenti in cellule trasfettate con il microRNA sintetico a confronto con cellule di controllo.

D.1 Selezione del microRNA 204-5p e prove preliminari

D.1.1 Selezione del microRNA

Il confronto di microRNA che da letteratura sono stati ritrovati down-regolati in fenotipi tumorali e predetti come potenzialmente capaci di legarsi alla 3'UTR della mesotelina, finalizzato all'individuazione di uno a comune, non ha fornito alcuna indicazione (fig. D.1 nella pagina successiva).

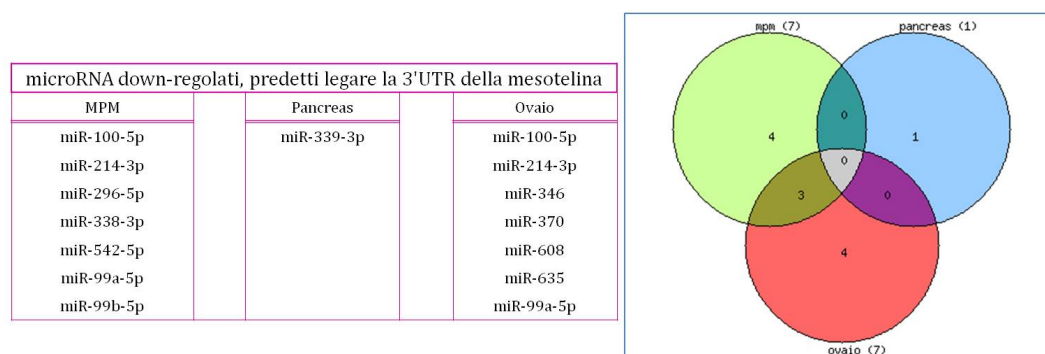


Figura D.1: In figura è riportato il risultato della comparazione tra i microRNA down-regolati nei tre tipi tumorali e predetti legare la 3'UTR della mesotelina

Si è tolto come filtro di selezione la predizione di legame, lasciando come unico requisito l'essere down-regolati in contesto tumorale(fig. D.2).

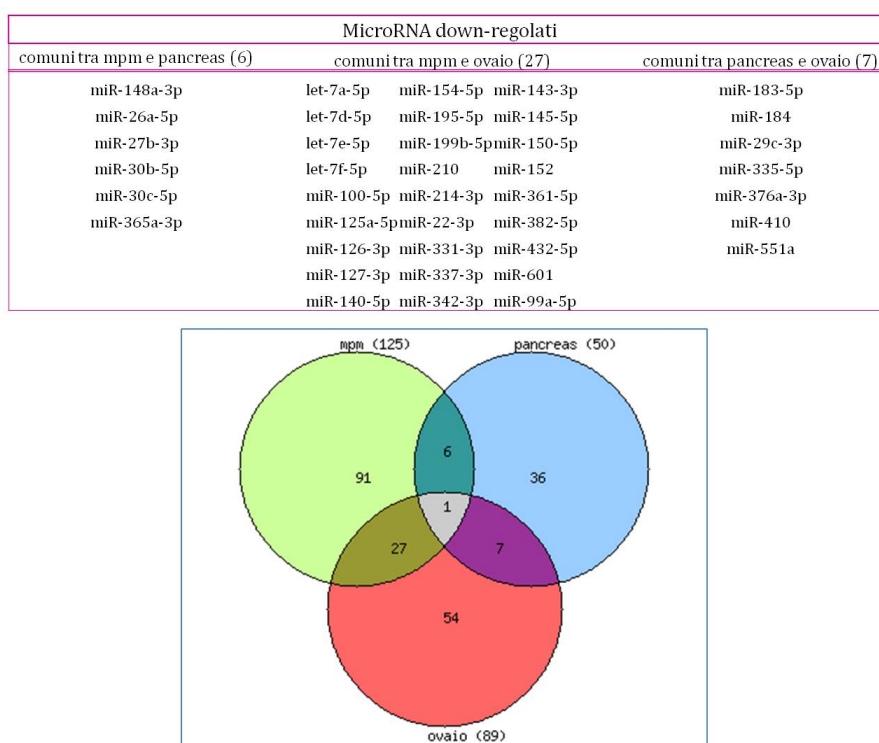


Figura D.2: In figura è riportato il risultato della comparazione tra i microRNA trovati in letteratura down-regolati nei tre tipi di tumore

Nell'immagine sono riportate le sequenze, sul messaggero della mesotelina, a cui il microRNA 204-5p è predetto legarsi con le tre più alte energie di legame (fig. D.3); in giallo, è messo in evidenza il sito complementare al seed del microRNA.

Figura D.3: Sono riportati, sulla sequenza del cDNA, i siti di nucleazione predetti da jMiRanda.jar con energie di legame maggiore.

D.2 Quantificazione allo spettrofotometro dell'RNA estratto nelle 5 prove sperimentali indipendenti

Nel complesso, le estrazioni sono risultate molto efficienti, tuttavia i rapporti di assorbanza 260/280 sono spesso risultati sotto l'ottimale (tranne che nell'ultimo esperimento, in cui il risultato è poco comprensibile).

Più che ad una contaminazione da proteine, i rapporti sub-ottimali potrebbero essere giustificati da residui di fenolo che non si è riusciti a rimuovere dopo la separazione di fase, dal momento che anche questo viene letto ad una lunghezza d'onda di 280nm.

É significativo notare poi come la resa delle estrazioni aumenti con i tempi sperimentali, da più bassa alle 24h, a più alta alle 48 e 72h; questo è perfettamente relazionato alla replicazione delle cellule nelle piastre, che da 80000 per p60 al momento della semina, arrivano a confluenza quasi completa alle 48/72h.

É riportato il risultato della corsa su gel degli RNA estratti nel primo esperimento (fig. D.4).

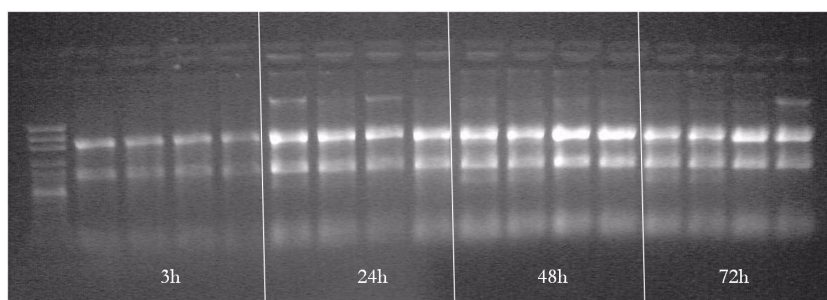


Figura D.4: Risultato della corsa elettroforetica degli RNA estratti nel primo esperimento

Per quanto non perfettamente integri, è possibile individuare le bande del 23S e del 28S, e riscontrare come l'intensità dell'estratto aumenti dalle 3h alle 48 e 72h.

Di seguito sono riportati i risultati delle quantificazioni allo spettrofotometro.

Esperimento	Tempo	Campione	Spettrofotometro		Quantitativi per RT	
			[mRNA](ng/μl)	A260/280	μl mRNA	H ₂ O
1	3h	miR A	343	1,738	2,92	12,08
		miR B	291	1,733	3,47	11,53
		CN A	335	1,382	2,99	12,01
		CN B	273	1,667	3,66	11,34
	24h	miR A	617	1,723	1,62	13,38
		miR B	575	1,664	1,74	13,26
		CN A	595	1,641	1,68	13,32
		CN B	650	1,542	1,54	13,46
	48h	miR A	805	1,657	1,24	13,76
		miR B	700	1,773	1,43	13,57
		CN A	1011	1,695	0,99	14,01
		CN B	899	1,617	1,11	13,89
	72h	miR A	876	1,569	1,14	13,86
		miR B	981	1,575	1,02	13,98
		CN A	1319	1,602	0,76	14,24
		CN B	1394	1,614	0,72	14,28
2	24h	miR A	629	1,663	1,6	13,4
		miR B	517	1,647	1,93	13,07
		miR C	682	1,662	1,47	13,53
		CN A	718	1,55	1,39	13,61
		CN B	714	1,599	1,4	13,6
		CNC	792	1,577	1,26	13,74
	48h	miR A	881	1,601	1,14	13,86
		miR B	876	1,609	1,14	13,86
		miR C	896	1,531	1,12	13,88
		CN A	773	1,555	1,29	13,71
		CN B	1358	1,47	0,74	14,26
		CNC	1033	1,611	0,97	14,03
	72h	miR A	778	1,404	1,28	13,72
		miR B	508	1,506	1,97	13,03
		miR C	672	1,5	1,49	13,51
		CN A	1026	1,553	0,97	14,03
		CN B	962	1,497	1,04	13,96
		CNC	1149	1,557	0,87	14,13

Tabella D.1: In tabella sono riportati i valori di lettura allo spettrofotometro degli estratti di RNA per gli esperimenti indipendenti 1 e 2.

Esperimento	Tempo	Campione	Spettrofotometro		Quantitativi per RT	
			[mRNA](ng/μl)	A260/280	μl mRNA	H ₂ O
3	24h	miR A	392	1,918	2,55	12,45
		miR B	314	1,806	3,18	11,82
		miR C	288	1,943	3,47	11,53
		CN A	518	1,637	1,93	13,07
		CN B	904	1,516	1,11	13,89
		CNC	675	1,346	1,48	13,52
	48h	miR A	879	1,474	1,14	13,86
		miR B	686	1,551	1,46	13,54
		miR C	501	1,504	2	13
		CN A	728	1,453	1,37	13,63
		CN B	874	1,351	1,14	13,86
		CNC	778	1,455	1,29	13,71
	72h	miR A	756	1,517	1,32	13,68
		miR B	795	1,472	1,26	13,74
		miR C	893	1,375	1,12	13,88
		CN A	1022	1,362	0,98	14,02
		CN B	1000	1,451	1	14
		CNC	1058	1,494	0,95	14,05
4	24h	miR A	549	1,431	1,82	13,18
		miR B	549	1,496	1,82	13,18
		miR C	627	1,427	1,59	13,41
		CN A	706	1,385	1,42	13,58
		CN B	650	1,389	1,54	13,46
		CNC	594	1,395	1,68	13,32
	48h	miR A	958	1,455	1,04	13,96
		miR B	829	1,518	1,2	13,8
		miR C	904	1,475	1,11	13,89
		CN A	1002	1,432	0,99	14,01
		CN B	829	1,473	1,21	13,79
		CN C	1061	1,414	0,94	14,06
5	24h	miR A	288	1,815	3,47	11,53
		miR B	342	1,584	2,93	12,07
		miR C	426	2,141	2,35	12,65
		CN A	134	1,941	7,16	7,54
		CN B	540	1,949	1,85	13,15
		CN C	462	2,26	2,16	12,84
	48h	miR A	664	2,155	1,51	13,49
		miR B	706	2	1,42	13,58
		miR C	624	2,124	1,6	13,4
		CN A	731	2,27	1,37	13,63
		CN B	834	2,113	1,2	13,8
		CN C	820	1,94	1,22	13,78

Tabella D.2: In tabella sono riportati i valori di lettura allo spettrofotometro degli estratti di RNA per gli esperimenti indipendenti 3,4 e 5.

D.3 Analisi statistica dei livelli di espressione del messaggero della mesotelina

E' riportata la tabella relativa all'ANOVA2 dei least square means con intervallo di confidenza al 95% per i livelli di espressione del messaggero della mesotelina normalizzati secondo lo Z-score.

Table of Least Squares Means for msln_norm_Zscore with 95% Confidence Intervals					
Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	54	-0,1507			
tempo					
24h	22	-0,4584	0,1463	-0,7526	-0,1643
48h	22	0,5868	0,1463	0,2926	0,881
72h	10	-0,5803	0,217	-1,0166	-0,144
miR					
0	27	0,0698	0,1413	-0,2143	0,354
1	27	-0,3711	0,1413	-0,6553	-0,087
tempo by miR					
24; 0	11	-0,3368	0,2069	-0,7528	0,0792
24; 1	11	-0,5801	0,2069	-0,9961	-0,1641
48; 0	11	1,0843	0,2069	0,6683	1,5003
48; 1	11	0,0892	0,2069	-0,3267	0,05053
72; 0	5	-0,538	0,3069	-1,1551	0,079
72; 1	5	-0,6226	0,3069	-1,2396	-0,0056

Tabella D.3: Tabella dei least square means con intervallo di confidenza del 95%

D.4 Analisi dei livelli di espressione proteica

D.4.1 Quantificazione degli estratti proteici per il modello 1 di Western Blot

Sono riportati i risultati delle letture degli standard (tabella D.4 nella pagina successiva), e la retta di taratura interpolata a partire da questi; le concentrazioni proteiche sono state ricavate risolvendo l'equazione della retta:

$$[proteina\ nel\ campione\ x] = \frac{(valore\ di\ assorbanza - b)/a}{3}$$

ricavando i µl di estratto da caricare per il western blot (riportato in tabella D.5 nella pagina seguente).

	$\mu\text{lBSA } 0,1\mu\text{g}/\mu\text{l}$	$[\text{BSA}](\mu\text{g}/\mu\text{l})$	$\text{H}_2\text{O a } 500\mu\text{l}$	$\text{SDS } 1\%(\mu\text{l})$	assorbanza	media	media - b1
b1					0 0	0	
1a	5	0,5	492	3	0,043	0,04	0,04
1b			492	3	0,037		
2a	10	1	487	3	0,07	0,067	0,067
2b			487	3	0,064		
3a	20	2	477	3	0,12	0,1155	0,1155
3b			477	3	0,111		
4a	50	5	447	3	0,24	0,246	0,246
4b			447	3	0,252		
5a	100	10	397	3	0,42	0,4275	0,4275
5b			397	3	0,435		
6a	150	15	347	3	0,608	0,6045	0,6045
6b			347	3	0,601		
7a	200	20	297	3	0,753	0,7515	0,7515
7b			297	3	0,75		
8a	400	40	97	3	1,038	1,0395	1,0395
8b			97	3	1,041		

Tabella D.4: Lettura allo spettrofotometro dell'assorbanza degli standard per la retta di taratura

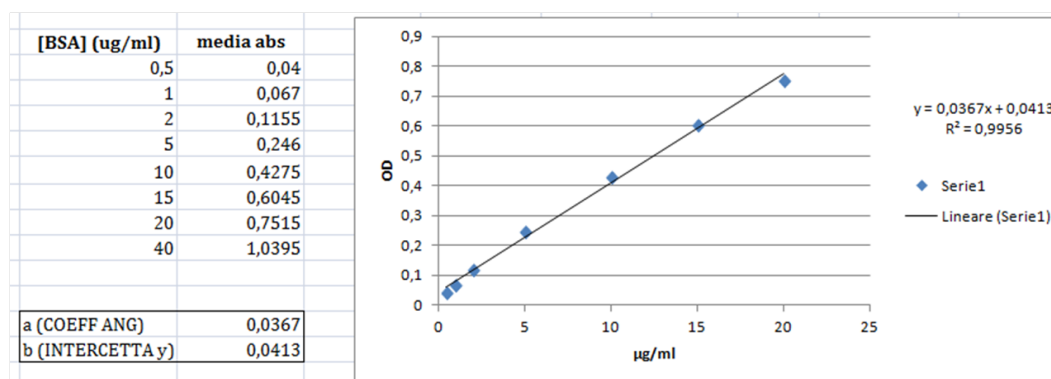


Figura D.5: Retta di taratura per la determinazione della concentrazione proteica dei campioni

Campione	Assorbanza	Media abs	Media - b1	concentrazione	$\mu\text{l per } 20\mu\text{g}$	$\mu\text{l SDS } 1\%$
miR A	0,24	0,245	0,245	1,8501	10,8100	2,1900
	0,25					
miR B	0,233	0,23	0,23	1,7139	11,6693	1,3307
	0,227					
miR C	0,256	0,242	0,242	1,8229	10,9716	2,0284
	0,228					
CN A	0,209	0,2145	0,2145	1,5731	12,7136	0,2864
	0,22					
CN B	0,219	0,2145	0,2145	1,5731	12,7136	0,2864
	0,21					
CN C	0,205	0,1935	0,1935	1,3824	14,4678	-1,4678
	0,182					

Tabella D.5: Lettura allo spettrofotometro dei campioni proteici estratti per l'esperimento 2, e loro quantizzazione

Quantificazione degli estratti proteici per il modello 2 di Western Blot

Nel secondo modello di western blot sono stati analizzati gli estratti proteici dei campioni dell'esperimento 2 alle 48h ed alle 72h dalla trasfezione; sono riportati i risultati della quantificazione delle proteine per alcuni campioni scelti in modo casuale tra le repliche sperimentali. Per la retta di taratura è stata utilizzata quella del modello 1 di western blot.

Campione	Assorbanza	Media abs	Media - b1	concentrazione	µl per 20µg	µl SDS1%
miR C (48h)	0,217 0,223	0,22	0,22	1,6231	12,3223	0,6777
CN B (48h)	0,283 0,297	0,29	0,29	2,2589	8,8540	4,1460
miR A (72h)	0,289 0,274	0,2815	0,2815	2,1817	9,1674	3,8326
miR B (72h)	0,291 0,271	0,281	0,281	2,1771	9,1865	3,8135
CN B (72h)	0,277 0,26	0,2685	0,2685	2,0636	9,6919	3,3081

Tabella D.6: Lettura allo spettrofotometro di campioni proteici selezionati tra gli estratti per l'esperimento 2, e loro quantizzazione

Risultati dei principali passaggi per l'analisi comparata in silico del gene *MSLN*

Sono di seguito riportati i risultati dei principali passaggi di analisi comparata della proteina e del cDNA della mesotelina, mirata a valutare lo stato di conservazione della sequenza codificante e quindi a complementare le indicazioni tratte dall'analisi sul microRNA, che è partita dal presupposto di un legame tra microRNA e messaggero a livello di sequenza codificante. Nel contempo, si cerca di individuare se nel corso della sua evoluzione il gene è andato ad acquisire –in particolari linee evolutive– caratteristiche peculiari in termini di sostituzioni aminoacidiche o di regioni o siti funzionali che ne hanno possibilmente modificato la funzione o la struttura.

E.1 Generazione degli allineamenti

Il primo passaggio richiesto dall'analisi evolutiva, una volta scelte le specie, è quello di allineare le sequenze proteiche e del cDNA, così che poi queste possano essere comparate tra loro.

Traduzione in silico mediante il Translator Tool di ExPASy

Viene riportato, a titolo esemplificativo, il risultato della traduzione in silico del cDNA di *Homo sapiens* (NM_005823.5), mettendo in evidenza le possibili cornici di lettura.

5'3' Frame 1

C Q A L H P H F P I E E T E A E E A Q R H A L L F L P
 G R P L P S A V T R G Q R A T G G P T V P P S L G S
 T Q T **Met** A L P T A R P L L G S C G T P A L G S L L F
 L L F S L G W V Q P S R T L A G E T G Q E A A P L D G
 V L A N P P N I S S L S P R Q L L G F P C A E V S G L
 S T E R V R E L A V A L A Q K N V K L S T E Q L R C L
 A H R L S E P P E D L D A L P L D L L F L N P D A F
 S G P Q A C T R F F S R I T K A N V D L L P R G A P E
 R Q R L L P A A L A C W G V R G S L L S E A D V R A L
 G G L A C D L P G R F V A E S A E V L L P R L V S C P
 G P L D Q D Q Q E A A R A A L Q G G G P P Y G P P S T
 W S V S T **Met** D A L R G L L P V L G Q P I I R S I P Q
 G I V A A W R Q R S S R D P S W R Q P E R T I L R P R
 F R R E V E K T A C P S G K K A R E I D E S L I F Y K
 K W E L E A C V D A A L L A T Q **Met** D R V N A I P F T
 Y E Q L D V L K H K L D E L Y P Q G Y P E S V I Q H L
 G Y L F L K **Met** S P E D I R K W N V T S L E T L K A L
 L E V N K G H E **Met** S P Q V A T L I D R F V K G R G Q
 L D K D T L D T L T A F Y P G Y L C S L S P E E L S S
 V P P S S I W A V R P Q D L D T C D P R Q L D V L Y P
 K A R L A F Q N **Met** N G S E Y F V K I Q S F L G G A P
 T E D L K A L S Q Q N V S **Met** D L A T F **Met** K L R T D
 A V L P L T V A E V Q K L L G P H V E G L K A E E R H
 R P V R D W I L R Q R Q D D L D T L G L G L Q G G I P
 N G Y L V L D L S **Met** Q E A L S G T P C L L G P G P V
 L T V L A L L L A S T L A **Stop** G P H S L A G P S P A
 G D P R L A R S R H G W S P F H P K R T R A Q **Stop** T
 G T C P L Q T R K K K K K K

5'3' Frame 2

A R L S T P T S Q L R K P R Q R R L S A T H S S F C L A
 G H S R L L **Stop** R A D R E L P V D P R C L P P W D L H
 R P W P C Q R L D P C W G P V G P P P S A A S C S C S S
 A S D G C S P R G P W L E R Q G R R L R P W T E S W P T
 H L T F P A S P L A N S L A S R V R R C P A **Stop** A R S
 V S G S W L W P W H R R **Met** S S S Q Q S S C A V W L T G
 S L S P P R T W T P S H W T C C Y S S T Q **Met** R S R G P
 R P A P V S S P A S R R P **Met** W T C S R G G L P S D S G
 C C L R L W P A G V C G G L C **Stop** A R L **Met** C G L W E
 A W L A T C L G A L W P S R P K C C Y P G W **Stop** A A R
 D P W T R T S R R Q P G R L C R A G D P P T A P R R H G
 L S P R W T L C G A C C P C W A S P S S A A S R R A S W
 P R G G N A P L G T H P G G S L N G P S S G R G S G G K
 W R R Q P V L Q A R R P A R **Stop** T R A S S S T R S G S
 W K P A W **Met** R P C W P P R W T A **Stop** T P S P S P T S
 S W T S **Stop** S I N W **Met** S S T H K V T P S L **Stop** S S T

W A T S S S R **Stop** A L R T F A S G Met **Stop** R P W R P
Stop R L C L K S T K G T K **Stop** V L R W P P **Stop** S T A
L **Stop** R E G A S **Stop** T K T P **Stop** T P **Stop** P P S T L G
T C A P S A P R S **Stop** A P C P P A A S G R S G P R T W
T R V T Q G S W T S S I P R P A L L S R T **Stop** T G P N
T S **Stop** R S S P S W V G P P R R I **Stop** R R S V S R Met
Stop A W T W P R S **Stop** S C G R **Met R C C R Stop** L W L
R C R N F W D P T W R A **Stop** R R R S G T A R C G T G S
Y G S G R T T W T R W G W G Y R A A S P T A T W S **Stop**
T S A C K R P S R G R P A S **Stop** D L D L F S P S W H C
S **Stop** P P P W P E G P T P L L A P A L L G I P A W P G
A G T G G P R S T P R E L A L S K R E H A P C R H V K K
K K K

5'3' Frame 3

P G S P P P L P N **Stop** G N R G R G G S A P R T P L S A
W P A T P V C C D A R T E S Y R W T H G A S L P G I Y T
D H G L A N G S T P V G V L W D P R P R Q P P V P A L Q
P R **Met G A A L E D P G W R D R A G G C A P G R S P G Q**
P T Stop H F Q P L P S P T P W L P V C G G V R P E H G
A C P G A G C G L G T E E C Q A L N R A A A L S G S P A
L **Stop** A P R G P G R P P I G P A A I P Q P R C V L G A
P G L H P F L L P H H E G Q C G P A P E G G S R A T A A
A A C G S G L L G C A G V S A E R G **Stop** C A G S G R P
G L R P A W A L C G R V G R S A A T P A G E L P G T P G
P G P A G G S Q G G S A G R G T P L R P P V D **Met V C L**
H D G R S A G P A A R A G P A H H P Q H P A G H R G R V
A A T L L S G P I L A A A Stop T D H P P A A V P A G S
G E D S L S F R Q E G P R D R R E P H L L Q E V G A G S
L R G C G P A G H P D G P R E R H P L H L R A A G R P K
A **Stop** T G **Stop** A L P T R L P R V C D P A P G L P L P Q
D E P **Stop** G H S Q V E C D V P G D P E G F A **Stop** S Q Q
R A R N E S S G G H P D R P L C E G K G P A R Q R H P R
H P D R L L P W V P V L P Q P R G A E L R A P Q Q H L G
G Q A P G P G H V **Stop** P K A A G R P L S Q G P P C F P
E H E R V R I L R E D P V L P G W G P H G G F E G A Q S
A E C E H G L G H V H E A A D G C G A A V D C G **Stop** G
A E T S G T P R G G P E G G G A A P P G A G L D P T A A
A G R P G H A G A G A T G R H P Q R L P G P R P Q H A R
G P L G D A L P P R T W T C S H R P G T A P S L H P G L
R A P L P C W P Q P C W G S P P G Q E Q A R V V P V P P
Q E N S R S V N G N **Met P P A D T Stop** K K K K K

Tra queste, è la prima sequenza (5'3' Frame1) quella che viene effettivamente trascritta:

5'3' Frame 1

tgccaggctctccacccccacttcccaattgaggaaccgaggcagaggaggtcagcgc
 C Q A L H P H F P I E E T E A E E A Q R
 cagcgaactcctcttctgcctggccggccactcccgtctgctgtgacgcgaggacagaga
 H A L L F L P G R P L P S A V T R G Q R
 gctaccggtggaccacaggtgcctccctcc|ctg|ggatctacacagacc|atg|gccttgcca
 A T G G P T V P P S | L | G S T Q T | M | A L P
 acggctcgacccctgttgggtctctgtggacccccgcctcggcagcctcctgttctctg
 T A R P L L G S C G T P A L G S L L F L
 ctcttcagcctc|g|ga|tgggtgcagccctcgaggaccctggctggagagacaggg|cag|gag
 L F S L | G | W V Q P S R T L A G E T G Q | E
 gctgcgccccgtggacggagtccctggccaaccacctaacatttcc|ag|c|ctctcccctcgc
 A A P L D G V L A N P P N I S S | L | S P R
 caactccttggttcccgtgtgcggaggtgtccggcctgagcacggagcgtgtccgggag
 Q L L G F P C A E V S G L S T E R V R E
 ctggctgtggccttggcacagaagaatgtcaagctctcaacagag|cag|ctg|cgtgtctg
 L A V A L A Q K N V K L S T E Q | L | R C L
 gctcaccggctctctgagcccccgaggacctggacgcctcccattggacctgctgcta
 A H R L S E P P E D L D A L P L D L L L
 ttctc|aa|c|cagatgcgttctcggggccccaggcctgcacccgtttcttctcccgcac
 F L N | P | D A F S G P Q A C T R F F S R I
 acgaaggccaatgtggacctgctcccaggggggtcccgcagcagcagcggctgctgcct
 T K A N V D L L P R G A P E R Q R L L P
 gcggctctggcctc|tgg|gg|gtgcgggggtctctgctgagcagcgtgatgtgcgggt
 A A L A C W | G | V R G S L L S E A D V R A
 ctgggagcctggcttgcgacctgcctggcgctttgtggcagtcgcccgaagtgcgtg
 L G G L A C D L P G R F V A E S A E V L
 ctacccccgctggtgagctgcccgggacccctggaccaggaccagcaggagcagccagg
 L P R L V S C P G P L D Q D Q Q E A A R
 gcggctctgcaggcggggggacccccctac|gg|c|ccccgctgcacatggtctgtctccag
 A A L Q G G G P P Y G | P | P S T W S V S T
 atggacgctctgcggggcctgctgccgtgctggccagcccatcatccgcagcatcccg
 M D A L R G L L P V L G Q P I I R S I P
 cag|ggc|atcgtggccgctggcggaacgctcctctcggaaccatcctggcggcagcct
 Q | G | I V A A W R Q R S S R D P S W R Q P
 gaacggaccatcctccggcccggttccggcggaagtggag|a|ag|acagcctgtccttca
 E R T I L R P R F R R E V E | K | T A C P S
 ggcaagaaggcccgagatagacgagagcctcatcttacaagaagtgggagctggaa
 G K K A R E I D E S L I F Y K K W E L E
 gcctgcgtggatgcggccctgctggccaccagatggaccgctgaacccatccccttc
 A C V D A A L L A T Q M D R V N A I P F
 acctacgagcagctggacgtcctaagcataaactggat|gag|ctc|taccacaaggttac
 T Y E Q L D V L K H K L D E | L | Y P Q G Y
 cccgagctctgtgatccagcacctgggtacctcttctcctcaagatgagccctgaggacatt
 P E S V I Q H L G Y L F L K M S P E D I
 cgcaagtggaatgtgacgtccctggagaccctgaaggcttctgctgaagtcaacaaaggg
 R K W N V T S L E T L K A L L E V N K G
 caggaatgagtcct|cag|gtg|gccaccctgatcgaccgttctgtaagggaaggggccag
 H E M S P Q | V | A T L I D R F V K G R G Q

```

ctagacaaagacaccctagacaccctgaccgccttctaccctgggtacctgtgctccctc
  L D K D T L D T L T A F Y P G Y L C S L
agccccgaggagctgagctccgtgccccccagcagcatc|tg|g|gggtcaggccccaggac
  S P E E L S S V P P S S I W|A|V R P Q D
ctggacacgtgtgacccaaggcagctggacgtcctctatcccaaggcccgcttgctttc
  L D T C D P R Q L D V L Y P K A R L A F
cagaacatgaacgggtccgaatacttcgtgaagatccagtccttctctgggt|g|gg|gcccc
  Q N M N G S E Y F V K I Q S F L G|g|A P
acggaggat|ttgaaggcgtcagtcagcagaatgtgagcatggacttgccacgttcatg
  T E D L K A L S Q Q N V S M D L A T F M
aagctg|cggacggatg|cgtg|ccg|ttgactgtggctgaggtgcagaaacttctggga
  K L R T D A V L|P|L T V A E V Q K L L G
ccccacgtggagggcctgaaggcggagggagcggcaccgcccgggtgcgggactggatccta
  P H V E G L K A E E R H R P V R D W I L
cggcagcggcaggacgacctggacacgctggggctggggctacagggcggcatccccaac
  R Q R Q D D L D T L G L G L Q G G I P N
ggctacctggctcctagacctcagcatgcaa|g|ag|gcctctcggggacgcctgcctccta
  G Y L V L D L S M Q|E|A L S G T P C L L
ggacctggacctgttctcaccgtcctggcactgctcctagcctccaccctggcc|tga|ggg
  G P G P V L T V L A L L L A S T L A - G
ccccactcccttgcctggccccagccctgctggggatccccgcctggccaggagcaggcac
  P H S L A G P S P A G D P R L A R S R H
gggtgggtccccgttccacccaagagaactcgcgctcagtaaacgggaacatgccccctg
  G W S P F H P K R T R A Q - T G T C P L
cagacacgtataaaaaaaaaaaaaaaaaa
  Q T R K K K K K K

```

Ottenuta la corrispondenza codone-amminoacido, è possibile sfruttare la distinzione in esoni ed introni fornita da Blat per trovare le giunzioni esone-esone, e quindi definire la fase dei singoli esoni sulla sequenza proteica, in quello che il primo passaggio per l'analisi degli allineamenti.

E.1.1 Allineamenti proteici

É riportato il risultato dell'allineamento ottenuto con MUSCLE per le sequenze proteiche del blocco 1 e del blocco 2, con la relativa divisione in esoni ed identificazione delle fasi; per i risultati dell'allineamento relativo al cDNA, si rimanda all'appendice G a pagina 155.

BLOCCO 1

Splicing phase	1	0
Exon number	1 2	2 3
Homo_sapiens	MALPTARPLLGSCTPALG-SLLFLLFSL GWVQPSRTLAGETGQ EAAPLD	49
Pan_troglodytes	MALPTARPLLGSCTPALG-SLLFLLFSF GWVQPSRALAGETGQ EAAPLD	49
Nomascus_leucogenys	MALLTARSLLGSCGTPALS-SLLLLLFSL GWVQPSRALAGETGQ EAAPLD	49
Macaca_mulatta	MALPMARPLSGSCGTPALG-SLLFLLFSL GWVQPSRVLAGETRQ EAAPLD	48
Papio_anubis	MALPTARPLLGSCTGTSALG-SLLFLLFSL GWVQPSRALAGETRQ EAAPLD	49
Saimiri_boliviensis	MALPTARPLWRSCGSPALG-RLPFLLLSL GWVQPKAWAGETGQ EAAPLD	49
Mus_musculus	MALPTARPLLGSCTGSPICRSFLLLLLSL GWIPRLQTQTTKTSQ EAT-LL	49
Rattus_norvegicus	MALPTAQPLLGSCTGSPICRSFLLLLLSL GWLPLLQTQTTKTSQ EAA-LL	49
Cricetulus_griseus	MALPTARPLLGSCTGSPICSSLLLLLSL GWMPVLQLQATKTDQ EAVHLC	50
	*** *:.* ***: : :***: **: : :.* :.* *	
Splicing phase	2	
Exon number	3 4	
Homo_sapiens	GVLANPPNISS LSPRQLLGFPCEVSGLSTERVRELAVALAQKNVKLSTE	99
Pan_paniscus	GVLANPPNISS LSPRQLLGFPCEVSGLSTERVRELAVALAQKNVKLSAE	99
Nomascus_leucogenys	GVLANAPNIAS LSPRQLLGFRCAEVSGLSTERVQELAVALGQKNVKLSAE	99
Macaca_mulatta	GILTANAPDIAS LSPRQLLGFTCEVSGLSTELVQELAVALGQKNVKLSAE	98
Papio_anubis	GVLTNAPDIAS LSPRQLLGFTCEVSGLSTELVQELAVALGQKNVKLSAE	99
Saimiri_boliviensis	RVLANTRNIAS LSPGGLLGFTCAEVSALESVWELAVSLGQKNVTLAE	99
Mus_musculus	HAVNGAADFAS LPTGLFLGLTCEEVSDLSMEQAKGLAMAVRQKNITLRGH	99
Rattus_norvegicus	HAVTGTVDFAFAS LPTGLFLGLTCEEVSGLSMGHAKELAMAVRQKNIVLQVH	99
Cricetulus_griseus	DELIGNPDFAS LPAGLFLGLTCEEVSGLSMAHAQALAMAVKQKNITLRVN	100
	: . :*: *.. :*: * *** * . **: :*: * .	
Splicing phase	0	2
Exon number	4 5	5 6
Homo_sapiens	Q LRCLAHRLSE--PPEDLDALPLDLLFLN PDAFSGPQACTRFFSRITKA	147
Pan_paniscus	Q LRCLAHRLSE--PPEDLDALPLDLLFLN PDAFSGPQACTRFFSRVTKA	147
Nomascus_leucogenys	Q LRCLAHRLSE--PPKDLDALPLDLLFLN PDAFSGPRACTRFFSRVTRA	147
Macaca_mulatta	Q LRCLAHQLSE--PPEDLDALPLDLLFLN PDAFSGPQACTHFFSRVAKA	146
Papio_anubis	Q LRCLAHRLSE--PPEDLDALPLDLLFLN PDTFSGPQACTHFFSRVAKA	147
Saimiri_boliviensis	Q LRCLAHRLSE--PPEDLDALPPDLLFLN PAAFSGSQACTRFFSRVSKA	147
Mus_musculus	Q LRCLARRLPRLTDEELNALPLDLLFLN PAMFPGQACAHFFSLISKA	149
Rattus_norvegicus	Q LRCLARRLPKHLTNEELDALPLDLLFLN PAMFPGQACAHFFSLISKA	149
Cricetulus_griseus	Q LRCLARRLPKYLTNEELDALPLDLLFFN PAMFRGQACVHFFSLISKA	150
	* *****:*. . :*:*** *****:* * * * :*:*** :*: *	
Splicing phase	0	
Exon number	6 7	
Homo_sapiens	NVDLLPRGAPERQRLPAALACW GVRGSLLEADVRALGGLACDLPGRFV	197
Pan_paniscus	NVDLLPRGAPERQRLPAALACW GVRGSLLEADVRALGGLACDLPGRFV	197
Nomascus_leucogenys	NVDLLPREAPERQRLAALACR GVRGSLLEADVQALGGLACDLPGRFV	197
Macaca_mulatta	NVDLLPRGAPERQRLPGALTWCW GVRGSLLEADVRALGGLACDLPGRFV	196
Papio_anubis	NVDLLPRGAPERQRLPAALACW GVQGSLLSEADVQALGGLACDLPGRFV	197
Saimiri_boliviensis	NVDLLPRGAPERRLPAALACR GVRGALLSEADVRALGGLACDLPGRFV	197
Mus_musculus	NVDVLPRLSLERQRLMEALKCQ GVYGFQVSEADVRALGGLACDLPGEFV	199
Rattus_norvegicus	NVNVLPRLSLERQRLTGALKCQ GVYGFQVSETDARALGGLACDLPGEFV	199

```

Cricetulus_griseus      NIDLLPQRSLEQRLLPRALECQ|GVYSFQLSETDAKALGGIVCDLPGKYV 200
*:::*: : **:*** ** * ** . :*:*.:*****.*****.:*

Splicing phase
Exon number
Homo_sapiens            AESAEVLLPRLVSCPGPLDQDQQEAARAALQGGGPPYG|PPSTWSVSTMDA 247
Pan_paniscus            AESAEVLLPRLVSCGLGPLDQDQQEAARAALQGGGPPYG|PPSTWSVSTLDA 247
Nomascus_leucogenys    AESAEVLLPRLARCLGPLDQDQQEAARAALQGGGPPYG|PPSTWSVSTLDA 247
Macaca_mulatta         AESAEVLLPRLVRCGLGPLDQDQQEAARAALQGGGPPYG|PPSTWSISTLDD 246
Papio_anubis           AESAEVLLPRLVRCGLGPLDQDQQEAARAALQGGGPPYG|PPSTWSISTLDA 247
Saimiri_boliviensis    AESAEVLLPRLVHCLGPLDQDQQEAARAALQGGGPPYG|PPSTWSASTLDA 247
Mus_musculus           ARSSEVLLPWLACGQGPLDQSQEKAVREVLRSGRTPYG|PPSKWSVSTLDA 249
Rattus_norvegicus      AKSSEVLLPWLARCGGPLDQQAQAVREVLRSGRAPYG|PPSTWSVSTLDA 249
Cricetulus_griseus     VKSSEVLLPWLACRGPLNEGQEEAVREALRSGRSPYG|PPSRWSVSTLNA 250
..*:.:*** *. * ***.: * :*. * .: * . ** *** ** *:.:

Splicing phase
Exon number
Homo_sapiens            LRGLLPVLGQPIIRSIPQ|GIVAARQRSSRDPSWRQPERTILRPRFRREV 297
Pan_paniscus            LRGLLPVLGQPIIRSIPQ|GIVAARQRSSRDPSWQPERTILWPRFRRDV 297
Nomascus_leucogenys    LQGLLPVLGRPIIRSIPQ|GMVARVAATLPWDPSWRQPEQTXLRPRFRRDV 297
Macaca_mulatta         LQSLLPVLGQPVISARPQ|GILAAWRQRSSRDPSWQPEQTVLRLRFRRDV 296
Papio_anubis           LQSLLPVLGQPVHSIPQ|GILAAWRQRSSRDPSWRQPEQTLRPRFRRDV 297
Saimiri_boliviensis    LRGLLPVLGQPVVRSIPQ|GVVAARKRTRSDPSWRQPERTAALRPRLRDMD 297
Mus_musculus           LQSLVAVLDESIVQSIKP|DVKAEWLQHSRDP SRLGSKLTVIHPFRFRDA 299
Rattus_norvegicus      LQGLLVLDDESIVHSIPK|DVITEWLQGISREPSRLGSKWTVTHPRFRRDT 299
Cricetulus_griseus     LQGLLVLDDESIIHSIPK|DVKAEWLQHISRGPSWRGIKLTAMIPFRFRDT 300
*:*.: **.....: : *: .: : . ** : * *:.:

Splicing phase
Exon number
Homo_sapiens            E|KTACPSGKKAREIDESLIFYKKWELEACVDAALLATQMDRVNAIPFTYE 347
Pan_paniscus            E|KTACPSGKKAREIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE 47
Nomascus_leucogenys    E|RTACPPGKEAHEINENLIFYKKWELEACVDAALLAQMDHVNNAIPFTYE 47
Macaca_mulatta         E|RTTCPPEKEVHEIDESLIFYKKWELEACVDPALLAQMDRVDAIPFTYE 346
Papio_anubis           E|RTTCPPEKEVHEIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE 347
Saimiri_boliviensis    E|KKACPPGKEVLEIDENLIFYKNWELEACVNASLLATQIDRVDAIPFTYE 347
Mus_musculus           E|QKACPPGKEPYKVDDELIFYQNWELEACVDGTMARQMDLVNEIPFTYE 349
Rattus_norvegicus      E|QKACPPGKEPNVVDENLIFYQNWELEACVDGTLLAGQMDLVNEIPFTYE 349
Cricetulus_griseus     E|RKACPPGMEPHKVDENLIFYQNWELEACVDGAMLATQMDLVNEIPFTYE 350
* .:.*. : :*:*****: *****: :*: * * : *****

Splicing phase
Exon number
Homo_sapiens            QLDVLKHKLDE|LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLTLKA 397
Pan_paniscus            QLDVLKHKLDE|LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLTLKA 397
Nomascus_leucogenys    QLDVLKHKLDE|LYPQGYPESVIQHDLHLFLKMSPEDIRKWNVTSLTLKA 397
Macaca_mulatta         QLDVLKHKLDE|LYPQGYPESVIRHLGHLFLKMSPEDIRKWNVTSLTLKA 396
Papio_anubis           QLDVLKHKLDE|LYPQGYPESVIRHLGHLFLKMSPEDIRKWNVTSLTLKA 397
Saimiri_boliviensis    QLGLKHKLDE|LYPQGYPEPVIQHLGQFFLKMSPEDIRKWNVTSLTLKA 397

```

Mus_musculus	QLSIFKHKLDK TYPQGYPESLIQQLGHFFRYVSPEDIHQWNVTSPTDVKT	399
Rattus_norvegicus	QLSIFKHKLDK TYPQGYPESLIKQLGHFFRYVSPEDIRQWNVTSPTDVNT	399
Cricetulus_griseus	QLDIFKHKLDE TYPQGYPESLIQQLGHFFRYISPEDIHRWNVSSPTDVKT	400
	.:*:*: *****.:*:*.* : * :*****:***:* :*:::	
Splicing phase	0	
Exon number	11 12	
Homo_sapiens	LLEVNGGHEMSPQ VATLIDRFVKGQQLDKDITLDTLAFYPGYLCSLSPE	447
Pan_paniscus	LLEVNGGHEMSPQ VATLIDRFVKGQQLDKDITLDTLAFYPGYLCSLSPE	447
Nomascus_leucogenys	LLEVSKGHKMSAQ VATVIDRFVKGQQLDKDITLDTLAFYPGYLCSLSPE	447
Macaca_mulatta	LLKVSKGHEMSAQ VATLIDRVVVGQQLDKDITVDTLTAFCPGCLCSLSPE	446
Papio_anubis	LLKVSKGHEMSAQ VATLIDRVVVGQQLDKDITVDTLTAFCPGCLCSLSPE	447
Saimiri_boliviensis	LLKVSKGHEMSVQ VATLIDHFLMGRQQLDKDITLDTLAFYPGYLCSLSPE	447
Mus_musculus	LLKVSQGQKMDAQ AIALVACYLRGGGQLEDMDVKALGDIPLSYLCDFSPQ	449
Rattus_norvegicus	LLKVSQGQKMDAQ VIALVACYLRGGGQLEDMDVKALDNIPLSYLCDFSPQ	449
Cricetulus_griseus	LLKVTKGQKMDAQ VTALVACYLRGGGHLDQDIVKAMDDIPLTYLCDFKPK	450
	.:*.*:*.* * . ::: : * :*:* :::: : **.:*.*:	
Splicing phase	2	
Exon number	12 13	
Homo_sapiens	ELSSVPPSSIW AVRPQDLDTCDPRQLDLVLYPKARLAFQNMNGSEYFVKIQ	497
Pan_paniscus	ELSSVPPSSIW AVRPQDLDTCDPRQLDLVLYPKARLAFQNMNGSEYFVKIQ	497
Nomascus_leucogenys	QLSSVPPSIIW AIRPQDLDTCDPRQLDLVLYPKARLAFQNMNGSEYFVKIQ	497
Macaca_mulatta	RLSSVPPSVIG AVRPQDLDTCDPRQLDLVLYPKARLAFQNMNGSEYFVKIR	496
Papio_anubis	RLSSMPPSIIW AVRPQDLDTCDPRQLDLVLYPKARLAFQNMNGSEYFVKIR	497
Saimiri_boliviensis	QLGSVPPSLIW AVRPQDLDTCDPRQLEVLVLYPKARLAFQNVNGSEYFSKIR	497
Mus_musculus	DLHSPVSSVMW LVGPQDLDKCSQRHLGLLYQKACSAFQNVSGLEYFEKIK	499
Rattus_norvegicus	DLHAIPSSVMW LVGLHDLDKCSQRHLGILYQKACSAFQNVSGLEYFEKIR	499
Cricetulus_griseus	NLHSIPSSFMW LVKPQDLNNSCSQRHLIDILYQKACSAFQNVSSRQLEYFEKIR	500
	* :*:.* : : :*:*. *:* :** ** **** . *** **:	
Splicing phase	1 0	
Exon number	13 14 14 15	
Homo_sapiens	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	547
Pan_paniscus	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	547
Nomascus_leucogenys	SFL GGAPTEDLKALSQQNASMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	547
Macaca_mulatta	PFL GGAPTELVKALSQQNVSMDLATFMKLREAVL PLTVAEVQKLLGPHV	546
Papio_anubis	PFL GGAPTEDLKALSQQNVSMDLATFMKLREAVL PLTVAEVQKLLGPHV	547
Saimiri_boliviensis	SFL GGAPAECLKLSQRNVSMDLATFMKLREAVL PLTVAEVQKLLGPHV	547
Mus_musculus	TFL GGASVKDLRALSQHNVSMDIATFKRLQVDSL GLSVAEVQKLLGPNI	549
Rattus_norvegicus	TFL GGASREDLRALSQHNVSMDIATFKKLQVDALV GLSVAEVQKLLGPHI	549
Cricetulus_griseus	TFL GGASVEDLRALSQNLSDMLATFKKLQVDALV GLTVAEVQKLLGPHI	550
	.** ***. :::: ***:* ***:*** :*: :::: *:*****:***:~	
Splicing phase	1	
Exon number	15 16	
Homo_sapiens	EGLKAEERHRPVRDWILRQRQDDLTGLGLQGGIPNGYLVLDLSMQ EAL	597
Pan_paniscus	EGLKAEERHPPVRDWILRQRQDDLTGLGLQGGIPNGYLVLDLSVR EAL	597
Nomascus_leucogenys	EGLKAEERHSPVRDWILGQRQDDLDALGLGLQGGIPNGYLVLDLSVR EAL	597
Macaca_mulatta	EGLKVEEQHSPVRDWILKQRQDDLTGLGLQGGIPNGYLILDSVR EAL	596

```

Papio_anubis      EGLKVEEQHSPVRDWILKQRQDDLDLTLGLGLQGGIPNGYLILDLSVR|EAL 597
Saimiri_boliviensis EGLKAEERHSPVRDWILRQRQDDLDLTLGLGLQGGIPNGYLILDRGVR|EAL 597
Mus_musculus      VDLKTEEDKSPVRDWLFRQHQLDRLGLGLQGGIPNGYLVLDNFVR|EAF 599
Rattus_norvegicus GDLKTEEDKSPVRDWLFRQQQKDLDSLGLGLQGGIPNGYLILDFNVR|EAF 599
Cricetulus_griseus ADLKTEEDKSPVRDWLFQQRQEDLDRLGLGLQGGIPNGYLILDNLVR|EAF 600
      .**.* : *****: *.**.* *****:*.**.: **:

Homo_sapiens      SGTPCLLGPVPVLTVLALLLASTLA- 622
Pan_paniscus      SGTPCLLGTGPVLTVLALLLASTLA- 622
Nomascus_leucogenys SGTPCLLGPAVLTVLALLLASTLA- 622
Macaca_mulatta    SGTPCLLGPVPVLTVLALLLASTLA- 621
Papio_anubis      SGTPCLLGPVPVLTVLALLLASTLA- 622
Saimiri_boliviensis SGTPCLLGPVPVLTVLALLAPTAC 623
Mus_musculus      SSRASLLGPGFVLIWIPALLPALRLS 625
Rattus_norvegicus SSGAPLLGPGFVFAWIPALLSALRLS 625
Cricetulus_griseus SSGAPLLGPGFVIAWIPALLSAFILS 626
      *. . ***. *: .: ***.

```

BLOCCO 2

```

Splicing phase      1      0
Odobenus_rosmarus   MSLQTARPLGSCGSAHGNLLLLLSL|GWVFPsRAQAADSRL|----- 41
Mustela_putorius    MAMRTAQPAWGSWGTSAAGSLLLLLSL|GWAFFsRAQAADLRL|----- 41
Canis_lupus         MALQTAQPPLGSRWTSTRGSLLLLLLSL|GWLVPsRAQAADSRL|----- 41
Felis_catus         MALQAALPPLGSRGTPHCSLLLLLSL|GWLVPsRVRAADSRP|----- 41
Bos_taurus          MALQTAPPRLGSWGTPAHGSLLLLLCL|GWLVPsRAQAADVGL|----- 41
Sus_scrofa          MALQTARPHLGSRGTRTPRSLLLLLSL|GWVPPsRAQAADTGQ|----- 41
Equus_caballus      MALQTAQRPLGSCGTPHGSLLLLLSL|GWLPPPQAQAADTGL|----- 41
Homo_sapiens        MALPTARPLLGSCGTPALGSLFLFLFSL|GWVQPSRTLGETGQ|EAAPLDGVLANPPNISS 60
      **: :*    ** *: .**:**.* ** *. . *:

```

```

Splicing phase      2      0
Odobenus_rosmarus   -----
Mustela_putorius    -----
Canis_lupus         -----
Felis_catus         -----
Bos_taurus          -----
Sus_scrofa          -----
Equus_caballus      -----
Homo_sapiens        |LSPRQLLGFPsCAEVsGLSTERVRELAVALAQKNVKLSTEQ|LRCLAHRLSEPPEDLDALPL 120

```

```

Splicing phase      2      0
Odobenus_rosmarus   -----
Mustela_putorius    -----
Canis_lupus         -----
Felis_catus         -----

```

```

Bos_taurus      -----
Sus_scrofa      -----
Equus_caballus  -----
Homo_sapiens    DLLLFLN|PDAFSGPQACTRFFSRITKANVDLLPRGAPERQRLPAALACW|GVRGSLSEA 180

```

```

Splicing phase                                     2
Odobenus_rosmarus -----
Mustela_putorius -----
Canis_lupus    -----
Felis_catus    -----
Bos_taurus     -----
Sus_scrofa     -----
Equus_caballus -----
Homo_sapiens    DVRALGGLACDLPGRFVAESAIEVLLPRLVSCPGPLDQDQQAARAALQGGPPYG|PPSTW 240

```

```

Splicing phase                                     0 1(2)
Odobenus_rosmarus -----|GVMTPLWLGTSRDPSSWQPELTVVLLRDRRDTE|-K 77
Mustela_putorius -----|GVMTPLWLGTSRDLVSWWRPELSVVLSDRRDTE|SK 78
Canis_lupus    -----|GVMTPLWLGTSRDLTWQWPELTVILSDRRDTE|KK 78
Felis_catus    -----|GVMTPLWLGTSRDLVSWQPELTVLLPRDRDTE|KK 78
Bos_taurus     -----|ALLSSSVQRTSRDLVSWQPELTVLLPRARRGTE|KK 78
Sus_scrofa     -----|GVLPWLQRTSGDLVSWQPEPTAIRPRPQGGTE|KR 78
Equus_caballus -----|GFVTPWLQRPTRDPSWQPELTVVLQRAQRSTK|S(|)K
Homo_sapiens    SVSTMDALRGLLPVLGQPIIIRSIPIQ|GIVAARQRSSRDPSSWRQPETILRPRFRREVE|KT 300

... . * .: : * ** : : * : ..

```

```

Splicing phase                                     0
Odobenus_rosmarus VCPSEKHAHVIDENLVFYEWELEACVDGALLAAQMDRVNMIPFTYQQLDVFKRKLDE|FY 137
Mustela_putorius  VCPPSRKAHVIDENLVFYEWELEACVDGALVAAQMDRVNMVPFTYQQLDVFKRKLDE|FY 138
Canis_lupus       VCPPGQKAHVVDENLVFYEWELEACVDGALLEAQMDRVNMVPFTYQQLDVFKRKLDE|LY 138
Felis_catus       ACPPERRARVVDENLVFYEWELEACVDGALLAAQMDRVNMVPFTYQQLGVFKRKLDE|LY 138
Bos_taurus        ACPPGREAQVVDENLVFYEWELEACVDGALLATQMRQVNLIPFTYQQLHIFKRKLDE|FY 138
Sus_scrofa        DCPGGRKAQVVDENLVFYEDWELEACVDGALLADQMDQVNLIPFTYQLHVFKLKLDE|LY 138
Equus_caballus    ACPPEQKATVVDENLVFYEWELEACVDAALLDAQMDRVNAIPFTHQQLIFRCKLNQ|FY 138
Homo_sapiens      ACPSGKKAREIDESLIFYKKWELEACVDAALLATQMDRVNAIPFTYQLDVLKHKLDE|LY 360

** .:* : ** *.** : *****. ** : ** :***:** : : ** : *

```

```

Splicing phase                                     0
Odobenus_rosmarus PQGYPESLIQHLYFFLELTPADIHKWNVTSLETVKSLLQVSKGQTMDAQ|VAALIARYVA 197
Mustela_putorius  PQGYPESLTQNLGYFFLELSPADIRKWNVTSLTVKSLLQVSKGQKRDAQ|VAALIARYVA 198
Canis_lupus       PQGYPESLIQHLGYFFFEITPEDIHKWNVTSLETVKSLLKVSQGHKQVDAQ|VAALIARYVA 198
Felis_catus       PQGYPESLVRHLGYFFLELTPEDIHKWNVMSLETVKSLLVSKGHEMDAQ|VAALIARYLV 198
Bos_taurus        PQGYPQSLIERLSYFFLYVTPEDIHKWNVTSLETVKSLLKVSQGRGVDAQ|VAALIARYVG 198
Sus_scrofa        PQGYPTSLVQRLRYFFLSVTPEDICKWNVTSLETVKSLLXVSKGRGMDAQ|VAALIARYVG 198
Equus_caballus    PQGYPESLIRRLRYFFYAVTPEDIYKWNVTSPETIVKYLLKISEGRMDAQ|VAVLIARYLV 198
Homo_sapiens      PQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLTLKALLEVNKGHEMSPQ|VATLIDRFVK 420

***** *: ..* *: * : * ** ***** * : : * ** :.: : ..* **.* : :

```

```

Splicing phase                                     2

```



```

Odobenus_rosmarus  GGGQLDKASLDTLATFGPTYLCFLSPEQLSSVQPSVLW|AARPQDLEACRPLQMDVLYPRA 257
Mustela_putorius   GGGQLDKATLDMLATFRPAYLCFLSLEQLRSVQHSVLW|AAGPQDLDACHPLQMAVLYPRA 258
Canis_lupus        GGGQLDKATLDVLATFSPTYLCVLSPEQLGVSQHSVIW|VARAQDLDACRPPQMDVLYPKA 258
Felis_catus        GGGQLDKATLDTLAAFHPTYLCFLSPEQLESVHHSVVW|AARPPDLDACRPAQMDVLYRKA 258
Bos_taurus         GGGQLDKAALDTLAIFRPAYLCFLRPEQLDSVQLSVLW|MTTPQDLDACSPPMQMAVLYHKA 258
Sus_scrofa         GGGQLDGATLDILATFCPTYLCFLSPEQLSSIQLSVVR|VTRPLDLDTCSPQVAVLYLKA 258
Equus_caballus     GGGQMDKATLDTLAAFRPTYLCLLSPEQLDSMPHNVVW|AAGPQDLDTCRPPQMDALYSKA 258
Homo_sapiens       GRGQLDKDLDTLTAFYPGYLCSLSPEELSSVPPSSIW|AVRPQDLDTC DPRQLDVLYPKA 480
* *:.* :.* *: * * * * * *:.* *: . : . . *.:* * *: .** :.*

Splicing phase           1                               0
Odobenus_rosmarus  RAAFQNINGSEYFVKIKPYL|AGAPTEDLRALSRRKVNMDMATFKTLQTEAVL|PLTIAEVQ 317
Mustela_putorius   RTAFQTVSGPEYFSRIKPFL|GGASTEDLRALSSQKVNMDVATFQKLQTEAVL|PLTIAEVQ 318
Canis_lupus        RTAFQNLGSGSEYFERIKPYL|GGARTEDLRALSQQNVSMdTATFKTLQREAVL|PLTVAEVQ 318
Felis_catus        RLAFQNMMSGSEYFEKMKLYL|GGAPTEDLRALSQRNISMDLAAFRTLRPEAVV|PLTVAEVR 318
Bos_taurus         HMAFQNVSGSEYFTRIKPFL|GGASTEDLRVFTSQNISIDAAAFKKLTTEAVL|SLTVAEVQ 318
Sus_scrofa         RAAFQNMMSGSEYLARIKPYL|GGAPTEDLRALSQRNISLDIATFKKLQTEVLR|KLTIAEVQ 318
Equus_caballus     RIAFQNMMSGSEYFARIKPYL|GGAPTEDLQALSRQNIHMDMATFKKLRTTEAVL|PLTIAEVQ 318
Homo_sapiens       RLAFQNMMSGSEYFVKIQSFL|GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL|PLTVAEVQ 540
: **.:.*:**: :: :* .** *****:.: :: :* *:* .* :.: **:**:

Splicing phase                               1
Odobenus_rosmarus  NLLGPNLVGLKAEQESSPVRDWILRQRQDDLDLSGLGLHGGIPNGYLVMDLSFR|EALSGG 377
Mustela_putorius   NLLGPNLVGLKAEQERSPVRDWIVRQQQDDLDLSGLGLHGGIPNGYLVMDLRLR|EALSGG 378
Canis_lupus        NLLGANLAGLKAEQGNSPVRDWILRQRQDDLDRLGLGLRGGIPNGYLVMDLSFR|EALSGG 378
Felis_catus        NLLGPNLVGLKAAESSPVRDWISRQRQEDLDLSGLGLRGGIPNGYLVVDLSFR|EALSGG 378
Bos_taurus         KLLGPNLVGLKAETGNMFLRDWISRQSQEDLDRLGLGLVGGVPNGYLVLDLRGR|EASSGG 378
Sus_scrofa         KLLGPHLAGLKAEENSPVRDWISRQPQDDLDRLGLGLQGGIPNGYLVLDINIR|EAPWGG 378
Equus_caballus     KLLGPNLAGLKAEENSPVRDWILRQPQDDLNLSGLGLRGGIPNGYLVLDLSFR|EALSGG 378
Homo_sapiens       KLLGPHVEGLKAEERHRPVRDWILRQRQDDLDLTLGLGLQGGIPNGYLVLDLSMQ|EALSGT 600
:***.:.: **** :**** ** *.:*: ***** **.:*****:.*: : ** *

Odobenus_rosmarus  AR-LGPG--PVFTAIP TLL LALIPK 399
Mustela_putorius   ARLVGPR--TVLTAIP TLL LALIPK 401
Canis_lupus        ARLLRPG--PVFTAILALLLGLIPK 401
Felis_catus        ARLLRPG--PVLTAVP TLL WALVPN 401
Bos_taurus         PHRLGQGPVLTVTASLLLVLVLS 403
Sus_scrofa         SRPLRPLAPVLTLP SLL LALILK 403
Equus_caballus     PRLLGPG--PGLTMILTLLAEILN 401
Homo_sapiens       PCLLGPG--PVLTVLALLLASTLA- 622
. : . :* ** :
```

L'analisi comparata ha permesso di rilevare un errore nella divisione in esoni ed introni riportata sul database genomico dell'UCSC Genome Browser relativamente ad *Equus caballus*: in tutte le specie l'esone 9 (E3 per i Laurasiateri, in cui gli esoni 3–8 sono deleti) termina con una fase 1, tranne che per il cavallo.

Analizzando nel dettaglio la sequenza:

```
... ag GGCTTCGTGA CCCCTGGCT GCAACGCCCC ACCCGGGACC CGTCCTGGCA
GCAGCCTGAG CTGACTGTCTG TCCTCCAGAG GGCCAGCGG AGCACAAAGA Gtgag ...
```

Viene correttamente riconosciuto l'accettore di splicing dell'introne 2 (ag), ma il donatore dell'I3 presenta un doppietto tg al posto del canonico GT; osservando bene la sequenza, si può vedere che se l'ultimo nucleotide dell'E3 (G) viene considerato come il primo nucleotide dell'introne, viene ripristinata la sequenza consensus dei siti donatori di splicing. Allo stesso tempo, così facendo, l'E3 termina con una fase 1, e non con una fase 2 che poco si accorda all'intero contesto ricavato dall'analisi comparata.

Si può quindi affermare che l'organizzazione in esoni ed introni è conservata filogeneticamente nei Placentalia, con le fasi ed i siti accettori e donatori di splicing canonici che vengono mantenuti in tutte le linee filetiche prese in considerazione.

Un maggior grado di divergenza emerge quando si confronta l'allineamento tra *Homo sapiens* e l'outgroup, *Monodelphis domestica*:

```
Homo_sapiens      -----
Monodelphis_domestica  MFSLTWSSPEVSPAPAPITLPHSSPSSPYSPFPVRVQGRTRVKLPASGGALKRTGSHTLWL 60

                                0           1           0

Homo_sapiens      -----
Monodelphis_domestica  RTEKAGAGQLSRLAELQEGTAQ|MTSQQAMGSLLCGPGIFLLFLAL|GCLATPGTHL|CTTGP 120

                                1           2           0

Homo_sapiens      -----
Monodelphis_domestica  ID|EAAICASVKS|ISEDEVRAASSRDHPTCHVQLAQYACIP|RPWLQVWKEDFLTSLYHCLV 180

                                0

Homo_sapiens      -----
Monodelphis_domestica  ASGPPSALAPTYSTLLFSKIDLPILETSLERFSQQ|FSQALLSDQWAGVLFNGLWGRLGPI 240

                                2

Homo_sapiens      -----
Monodelphis_domestica  FQAREAQVTWAQWLSGIQPFLHPEAFYCLAAQNTSCETFHNL|ISVLNDVYSSLEAEKQR 300

                                1           0

Homo_sapiens      -----
Monodelphis_domestica  KIFEGIEAFLQE|GSHCQEETVPLNSSSWLANYLGGFFLERATEEELKKLVDEPK|FQMLA 360

                                2

Homo_sapiens      -----
Monodelphis_domestica  RDPATLHLVSQLHLTPDLASFYAALLTSPPRINFSS|IPDTLICHLRPEALGKIEFEESLA 420
```

```

0
Homo_sapiens -----
Monodelphis_domestica LVQRLNKACFDPPPQLDGNKSVSEPPSWEALQ|VASRLLSGFGDLSEAVLRSLGPAAVGLP 480

0
Homo_sapiens -----MALP-----TARPLL-----SCGTPALGSLFL 24
Monodelphis_domestica VSQIEEHLGGRDLVAALPALAQVRGWSRQARALLGKLLESGYE|VSDAQSLAALGSLVSG 540
*** **.*** :.*****:

1 0
Homo_sapiens LFSL|GWVQPS-RTLAGETG-----Q|----- 42
Monodelphis_domestica LSSA GLQSLAPRVVLEAMGEPGFAQHLAHIPALLKTAFFEQ|LALAVPNPITLVQAVPDTL 600
* * * . : * : * *

0 2
Homo_sapiens -----|EAAPLDGVLANPPNISS|LSPRQLLGFPCL 72
Monodelphis_domestica TEAIPSAILAFSLDDQPRFDDLNSRRWTAVQ|AAMFFNEVVKNVSDFDS|ISPYILHGFTCA 660
* : : * * : : * : * * * *

0 2
Homo_sapiens EVSGLSTERVRELAVALAQKNVKLSTEQ|LRCLAHRLSE---PPEDLDALPLDLLFLN|PD 129
Monodelphis_domestica SASSLDLEQVHRLAKVMERKNVTLGAEQ|LSCIAKRVSENGIPEALDEYPKDMLFLS|PS 720
..*.*.*:.*. : :***.*:* * ***:** * ** * :*****. *.

0
Homo_sapiens AFSGPQACTRFFSRITKANVDLLPRGAPERQRLLPALACW|GVRGSLLEADVRALGGLA 189
Monodelphis_domestica AYTRTGGCKHFFTRVGGQSNVDILQKDSLRSRLSEALTCL|GISDTRVTKEDICILGQLA 780
*: . *. :*: : :***: : : *.*. **: * : : : * : * *

2
Homo_sapiens CDLPGRFVAESAELVLLPRLVSCPGPLDQDQQAARAALQGGGPPYG|PPSTWSVSTMDALR 249
Monodelphis_domestica CDLDGRYINDSAEVLPLQLEQCGGPFSPDQKEATNMALRSGDSPYG|PPSRWSVSTLNALQ 840
*** **: :*****:* . * *: . **:*. :*.*.*** ** * :*: :*:

0 1
Homo_sapiens GLLPVLGQPIIRSIPQ|GIVAARQRSSRDPSWRQP-----ERTILRPRFRREVEK|TACPS 304
Monodelphis_domestica GLLPVLDRHIQINIPQ|SIVTSWLKRALLDPSWPRENRAFIQNLQSSRHRDASE CLEVS 900
*****.: **:*.*** .*: : * : * : : : : : : : : : : : : : : : : : : :

0 0
Homo_sapiens GKK AREIDESLIFYKKWELEACVDAALLATQMDRVNAIPFTYEQLDVLKHKLDE|LYPQGY 364
Monodelphis_domestica GSS |ITEISEELVIYTEKELWECLNTSLLAANLDKLTIPFTYQQEEVFKAFLDK|LYPDGY 960
*.. **.*.*:.*. : * : : : : : : : : : : : : : : : : : : : : : : :

0
Homo_sapiens PESVIQHLGYLFLKMSPEDIRKWNVTSLETLKALLEVNKGHEMSQ|VAT----- 413
Monodelphis_domestica PEVVNNLGSFLKLVTPEDIQKNVTSVETLTLQATDKNDMVT WLPNHEYISRLGWE 1020
** **:.* ** : :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*:

1 2
Homo_sapiens -- LIDRFVKGRGQLDKDTLDTLAFYPGYLCSLSPEELSSVPPSSIW|AVRPQDLDTCDPR 471
Monodelphis_domestica FW|VINRYIQGGGLNMAALNVIG---SKYFCSLTTEEQLNSIEPSEFR|MVNSLNPACSQA 1077
*: :*: :* * : : : . * :*: :*.*: * : * . : : : .

1
Homo_sapiens QLDVLYPKARLAFQNMNGSEYFVKIQSFL|GGAPTEDLKALSQQNVSMDLATFMKLRTDAV 531
Monodelphis_domestica QKKILYPKARTAFQGILFPQYRLIQPYL|DGAPPQDIKDSLQKNVNMIDIGTFKKLQKEAI 1137
* . :***** **: . : * :*.***.:* ***:**.*:.* ***: :*:

0
Homo_sapiens L|PLTVAEVQKLLGPHVEGLKAEERHRPVRDWILRQRQDDDLTLGLGLOGGIPNGYLVLDL 591

```

```

Monodelphis_domestica  L|PLTPPEVRGLLGQNLPLGLKAEKTSVPREWILKQRQDSLALGIGLTGGIPNGYIIISR 1197
* *** **: *** :: *****: ***:***:****.***:***:*** *****:..
1
Homo_sapiens           SMQ|EALSGTPCLLGPGPVLTVLALLLASTLA 622
Monodelphis_domestica  RNR HR----- 1202
: .

```

I primi 15 esoni riportati per l'opossum sono significativamente differenti rispetto a quanto non si riscontri nei placentati, dove sono quasi completamente assenti salvo per una minima omologia tra gli esoni 13–14 del marsupiale ed esoni 1–2 dei placentati. Per quanto riguarda invece l'organizzazione in introni ed esoni, il blocco costituito dagli esoni 16–21 (3–8) è perfettamente conservato, così come le fasi finali degli esoni 25–27 (12–14); in caso contrario, si sarebbe stati legittimati a dubitare dell'attribuzione del gene ENSMODG00000016813 quale ortologo della gene *MSLN* umano, venendo a mancare sufficienti supporti all'omologia di sequenza.

Un ulteriore elemento interessante che è emerso dall'analisi è la conservazione perfetta del codone di stop a livello dei placentati, TGA in tutti i taxa analizzati, ma differente rispetto al marsupiale, dove il codone è codificato dalla tripletta TAA.

E.2 Alberi filogenetici

Con il programma Mega5 sono stati generati gli alberi filogenetici relativi al blocco 1 ed al blocco 2 per la proteina ed il cDNA della mesotelina, utilizzando come modello il Maximum Likelihood; vengono mostrati alberi in cui le distanze sono quelle realmente calcolate, poiché la topologia è quella dell'albero migliore tra tutti quelli testati, ed in cui i valori di bootstrap sono stati aggiunti dal programma in un secondo momento.

Si riportano solo gli alberi più significativi tra tutti quelli generati, ottenuti includendo tutte le specie dei blocchi 1 e 2 (con e senza outgroup così da valutare le eventuali differenze) a livello della sequenza completa aminoacidica e del cDNA, in modo da avere una indicazione sulle relazioni evolutive del gene. Tutti gli alberi in cui i due blocchi sono stati analizzati indipendentemente sono riportati nell'appendice I a pagina 187.

Blocchi congiunti; sequenza proteica intera, senza outgroup (fig. E.1).

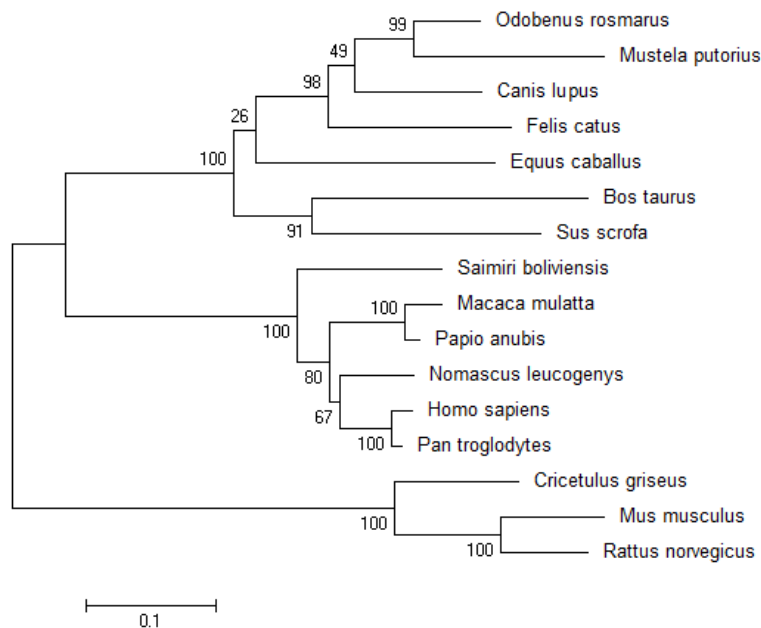


Figura E.1: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza proteica intera senza l'outgroup

Si può apprezzare come nel complesso tutti i nodi abbiano una buona affidabilità statistica, che porta a poter ritenere l'albero come una buona ricostruzione del processo evolutivo seguito dal gene per quanto riguarda la sequenza proteica. La topologia ricalca la filogenesi tradizionale delle specie, eccezion fatta per *Equus caballus*, i cui bassi valori di bootstrap indicano come non si riesca ad inferire correttamente e con precisione la sua collocazione. Essendo un albero senza radice, i Primates non clusterizzano con i Laurasiatheria come potrebbe sembrare dalla raffigurazione grafica, piuttosto si distinguono tre grandi cluster, quello dei Rodentia, dei Primates e dei Laurasiatheria.

Blocchi congiunti; sequenza proteica intera, con outgroup (fig. E.2 nella pagina seguente).

La maggior parte dei nodi è supportata da alti valori di bootstrap, e la storia evolutiva del gene ricalca quella delle specie; unica eccezione è rappresentata da *Equus caballus*, che tende a clusterizzare con i *Carnivora* come nel precedente albero (l'incertezza della collocazione è sottolineata da valori bootstrap inferiori alla significatività, 25%).

Poco solida, per quanto concorde con la filogenesi di riferimento è anche la separazione tra *Canis lupus familiaris* e gli Arctoidea (Musteloidea e Pinnipedia). Da un punto di vista della topologia, vengono distinti questa volta due cluster principali oltre all'out-group, quello degli Euarchontoglires, in cui si includono Primates e Rodentia, e quello dei Laurasitheria, come da accordo con le ricostruzioni filogenetiche dei Placentati su base molecolare.

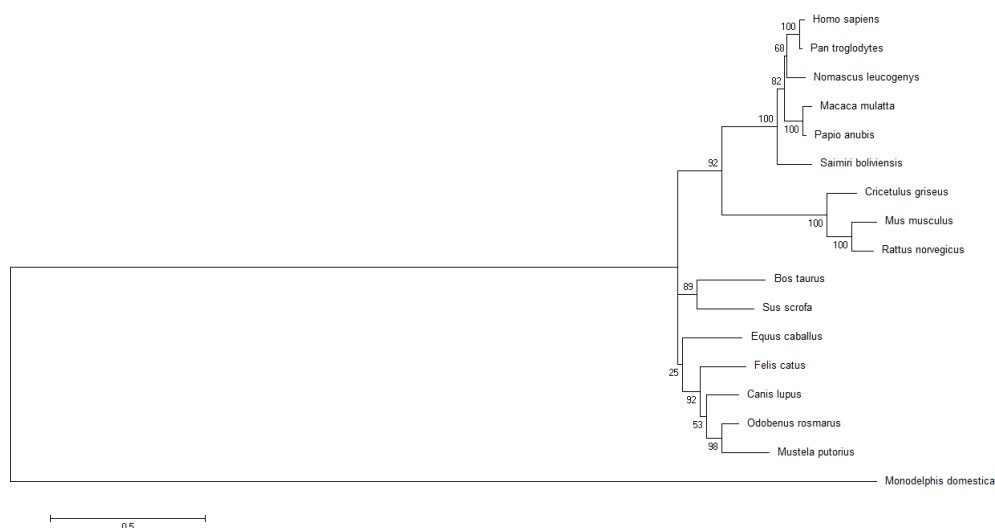


Figura E.2: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza proteica intera ed includendo l'outgroup

Blocchi congiunti; cDNA per intero, senza outgroup (fig. E.3 nella pagina successiva).

L'albero presenta nel complesso una alta significatività statistica delle diramazioni, anche nel caso di *Equus caballus*, che viene nuovamente collocato come prima linea di emergenza nel cluster dei Carnivora. Come nel caso della ricostruzione basata sulla sequenza proteica (E.1 nella pagina precedente), si distinguono tre grandi cluster, quello dei Rodentia, dei Primates e dei Laurasiatheria, senza che sia possibile collocare la radice per discriminare le specifiche relazioni.

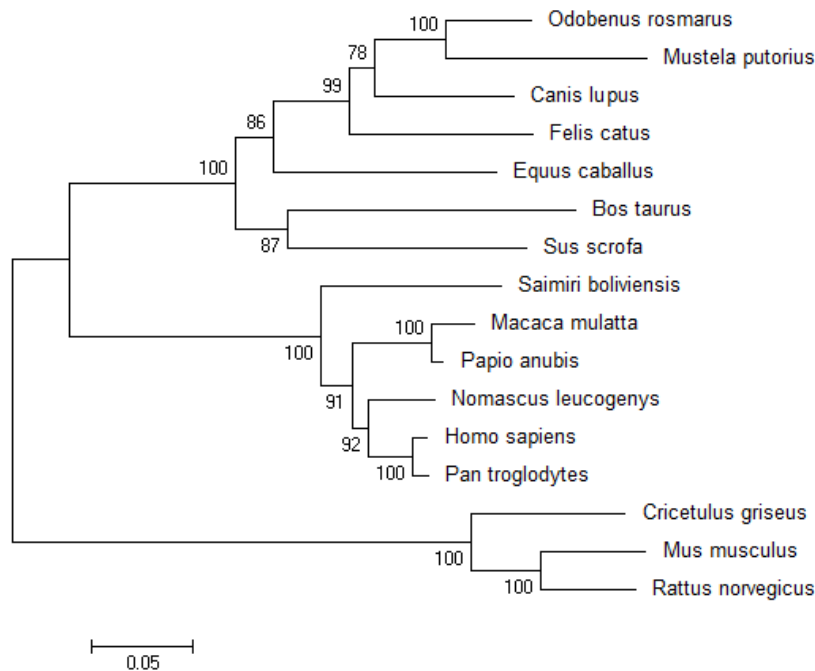


Figura E.3: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza del cDNA intera senza l'outgroup

Blocchi congiunti; cDNA per intero, con outgroup (fig. E.4 nella pagina successiva).

L'introduzione dell'outgroup non porta a cambiamenti nella ricostruzione filogenetica del gene, con nodi che registrano alti valori di bootstrap. A differenza dell'albero a blocchi congiunti basato sulla sequenza della proteina (E.2 a fronte con outgroup, dove i Rodentia costituivano il sister group dei Primates, qui rappresentano il sister group dei laurasiateri, mentre il cluster dei primati si pone come blocco a se stante, separato rispetto all'outgroup.

Essendo questa la ricostruzione più completa, che include tutte le informazioni a disposizione per il gene, si può ipotizzare che da un punto di vista del cDNA, il gene della mesotelina indichi una maggiore affinità tra roditori e laurasiateri piuttosto che tra roditori e primati. Per contro, la sequenza proteica è più concorde con la filogenesi di riferimento per le specie, ed i roditori sono più vicini ai primati che ai laurasiateri, con gli artiodattili che emergono alla base della diramazione di questi ultimi.

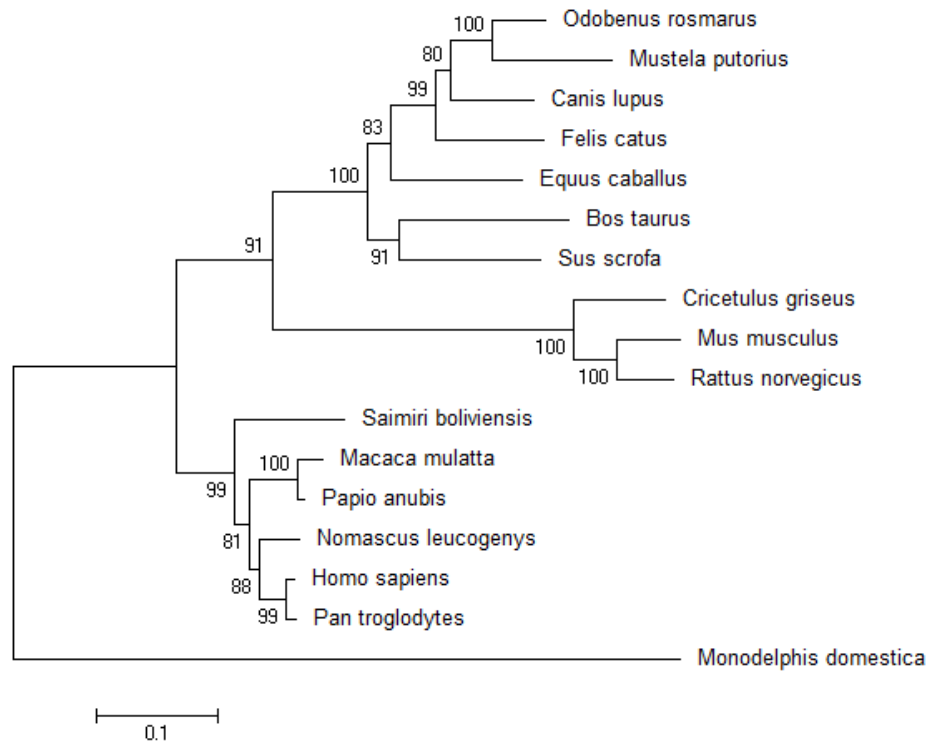


Figura E.4: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza del cDNA intera con l'outgroup

In un secondo momento, sono stati generati gli alberi funzionali allo studio con il programma PAML, ristretti alla porzione di sequenza compresa tra l'esone 9 e l'esone 16 (in riferimento alla sequenza del blocco 1), che è poi quella in cui vi è il precursore della mesotelina matura.

Blocchi congiunti; sequenza proteica E9–E16, senza outgroup (fig. E.5 nella pagina successiva).

La ricostruzione topologica è sostenuta da forti valori di bootstrap, fatta eccezione per i nodi deboli già precedentemente individuati che collocano *Equus caballus* come parte del cluster dei Carnivora (19%), e quello che porta *Canis lupus familiaris* a clusterizzare, come siser group, con gli altri Caniformia Arctoidea. Da un punto di vista della topologia si denotano ancora una volta tre cluster principali, Rodentia, Primates e Laurasiatheria, proprio come negli alberi senza outgroup del cDNA e proteina a sequenza intera (alberi in fig. E.1 a pagina 141 e E.5 nella pagina successiva).

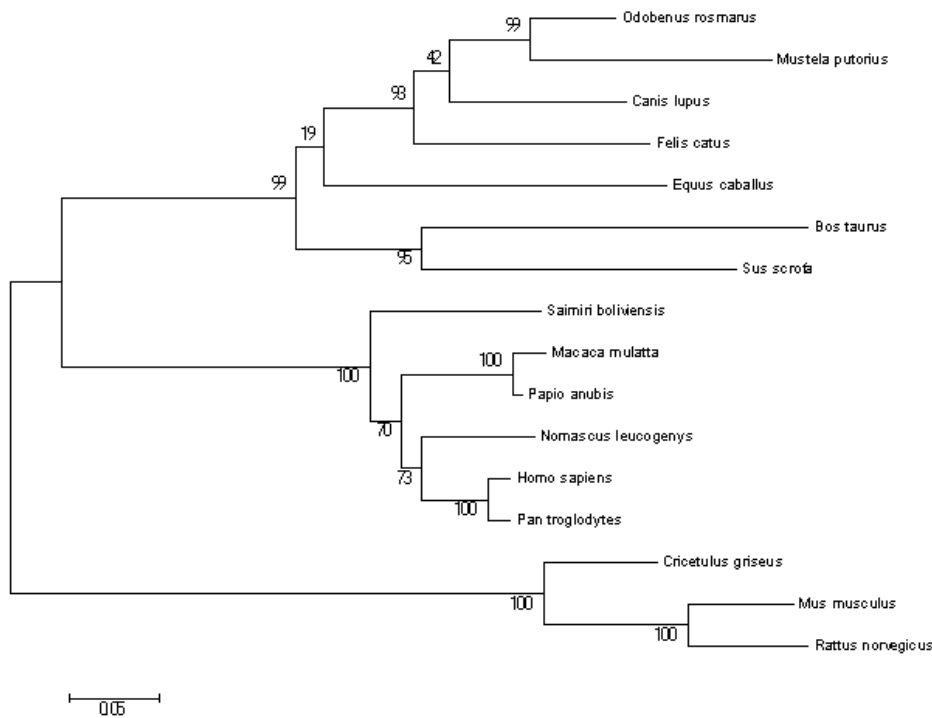


Figura E.5: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza proteica E9–E16 senza l’outgroup

Blocchi congiunti; sequenza proteica E9–E16, con outgroup (fig. E.6 nella pagina seguente).

L’introduzione dell’outgroup nell’allineamento determina una maggiore risoluzione generale dell’albero, anche se i due nodi più deboli rimangono anche in questo caso al di sotto della significatività. La sola porzione di sequenza codificante per la mesotelina matura non è tuttavia sufficiente a risolvere i nodi più ancestrali dell’albero, e si trovano infatti quattro cluster principali, uno rappresentato dal solo outgroup, uno dai Rodentia, un altro dai Primates (che a differenza della sequenza proteica intera, E.2 a pagina 142, non sono il sister group dei roditori) e l’ultimo dei Laurasiatheria. Se la radice viene posta alla base, tra marsupiali e placentati, allora la diramazione successiva ottenuta non è una dicotomia, ma una triplice diramazione che non si accorda alle regole formali della ricostruzione filogenetica.

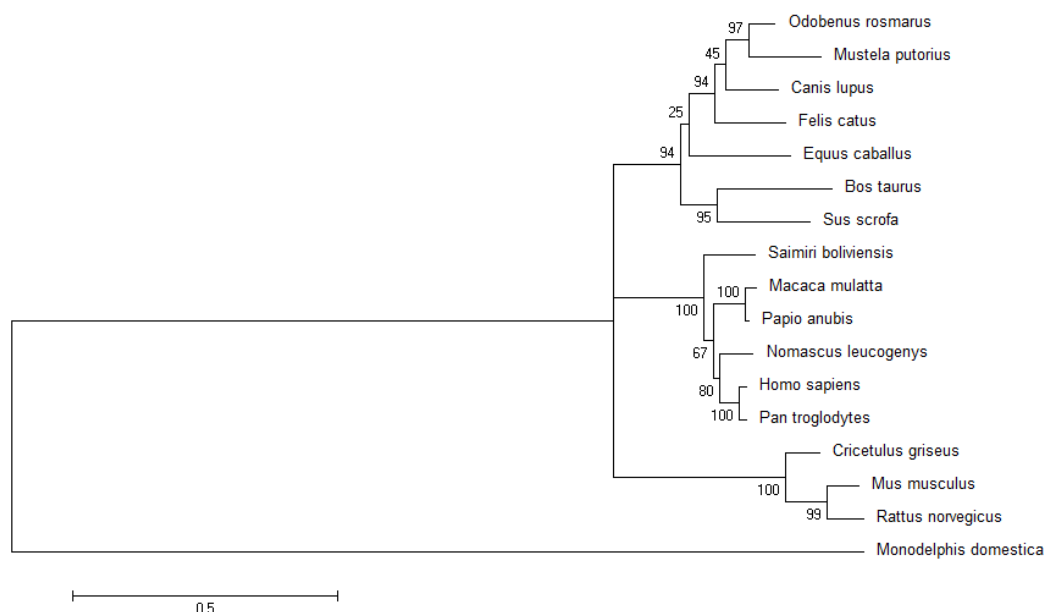


Figura E.6: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza proteica E9–E16, con l’outgroup

Blocchi congiunti; cDNA E9–E16, senza outgroup (fig. E.7 a fronte).

É da segnalare una imprecisione generata dal programma, ossia l’introduzione della radice alla base della divergenza tra il cluster dei roditori e di quello che raggruppa primati e laurasiateri, che porta a non poter stabilire con certezza se vengono riconosciuti due cluster principali o tre, come negli alberi precedenti a blocchi congiunti senza outgroup. A parte questo, anche i nodi deboli hanno qui una significatività statistica superiore al 50%, e confermano le indicazioni già ricavate.

Blocchi congiunti; cDNA E9–E16, con outgroup (fig. E.8 a pagina 148).

Tutti i nodi dell’albero sono supportati da valori di bootstrap significativi (>50%), e si può quindi ritenere la ricostruzione come indicativa del processo evolutivo seguito: si riconoscono tre cluster principali, quello dell’outgroup, quello dei Primates ed un terzo che racchiude da una parte i Rodentia e dall’altro i Laurasiatheria. La lunghezza dei rami suggerisce che tra tutte le specie analizzate, la più divergente è proprio l’outgroup (che può quindi considerarsi valido), seguita solo dal gruppo dei roditori, il cui ramo

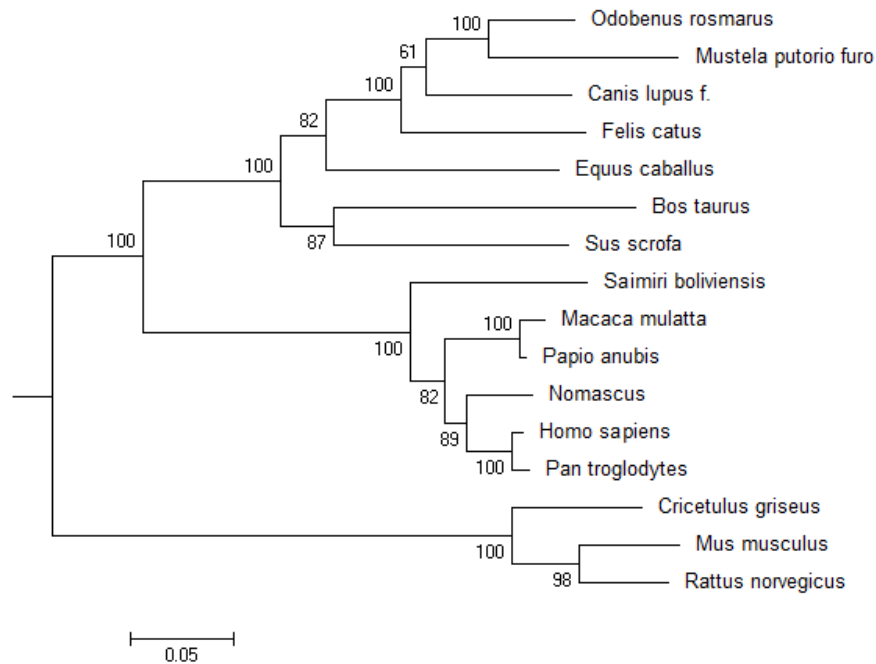


Figura E.7: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza di cDNA E9–E16, senza l’outgroup

che li separa dai laurasiateri ha una lunghezza superiore alla media. Si ha poi nuovamente riprova di come *Equus caballus* abbia per questo gene una maggiore affinità con i Carnivora che con gli Artiodactyla, di cui sono invece evolutivamente il sister-group in accordo con il record paleontologico e con la filogenesi molecolare delle specie.

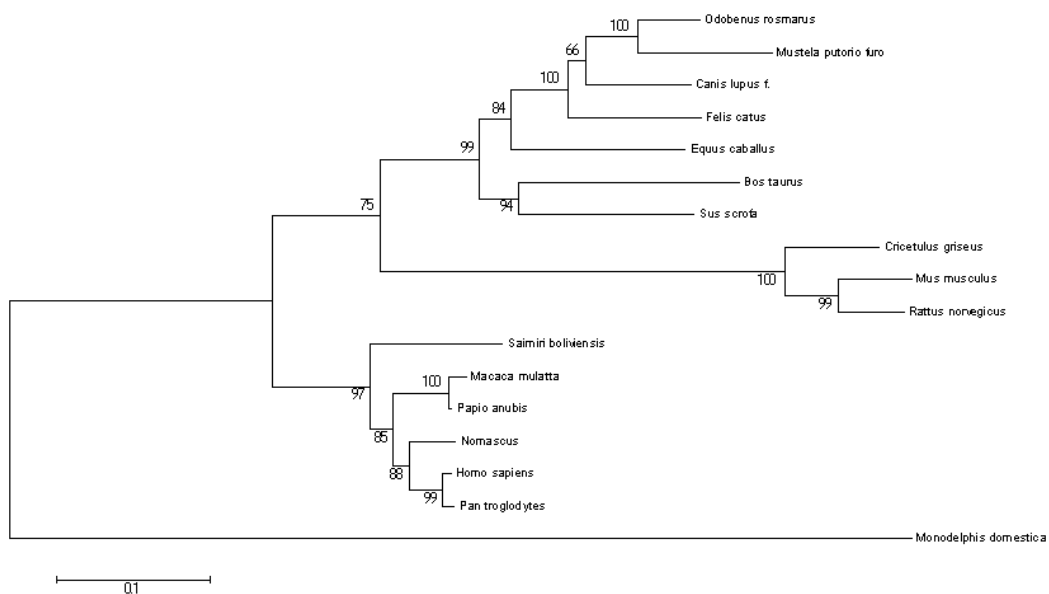


Figura E.8: Albero di maximum likelihood per le 16 specie di placentati, analizzando la sequenza del cDNA E9-E16, con l'outgroup



Risultati dell'analisi comparata del microRNA

miR-204-5p

Quando negli allineamenti multipli a blocchi sono stati evidenziati tutti i possibili siti di legame del microRNA, sono stati individuati i seguenti siti conservati per tutte le specie:

F.0.1 Blocco1

Esone	5	6
Homo	GGACCTGCTGCTATTCTCTCAA CCCAGATGCGTTCTCGGGGCCCCAGGCCTGCACCCGTTT	587
Pan	GGACCTGCTGCTATTCTCTCAA CCCAGATGCGTTCTCGGGGCCCCAGGCCTGCACCCGTTT	581
Nomascus	GGACCTGCTGCTATTCTCTCAA CCCAGACGCGTTCTCGGGGCCCCGGGCCTGCACCCGTTT	580
Macaca	GGACCTGCTGCTCTTCTCTCAA CCCAGACGCGTTCTCGGGGCCCCAGGCCTGCACCCACTT	416
Papio	GGACCTGCTGCTATTCTCTCAA CCCAGACGCGTTCTCGGGGCCCCAGGCCTGCACCCACTT	577
Saimiri	GGACCTGCTGCTCTTCTCTCAA CCCGGCCGCGTTCTCAGGGTCCCAGGCCTGCACCCGTTT	520
Mus	GGACCTGCTGCTCTTCTCTCAA CCCAGCCATGTTTCCAGGGCAACAGGCTTGTGCCACTT	533
Rattus	GGACCTGCTGCTCTTCTCTCAA TCCAGCCATGTTTCCGGGGCAACAGGCTTGTGCCACTT	515
Cricetulus	GGACCTGCTGCTCTTCTCTCAA CCCAGCCATGTTTCCAGGGCAACAAGCCTGTGTCCACTT	428
	***** .*** **	*** .*** ..** ** . **
Homo	CTTCTCCCGCATCACGA AGGCCAATGTGGACCTGCTCCCGAGGGGGG	CTCCCGAGCGACA 647
Pan	CTTCTCCCGCGTCACGAAGGCCAATGTGGACCTGCTCCCGAGGGGGGCTCCCGAGCGACA	641
Nomascus	CTTCTCCCGCGTAACGAGGGCCAACGTGGACCTGCTTCTCGGGAGGCTCCTGAGCGACA	640
Macaca	CTTCTCCCGCGTCGCGAAGGCCAACGTGGACCTGCTCCCGCGGGGGGCTCCTGAGAGACA	476
Papio	CTTCTCCCGCGTCGCGAAGGCCAACGTGGACCTGCTCCCGCGGGGGGCTCCTGAGAGACA	637
Saimiri	CTTCTCCCGCGTCTCAAAGGCCAATGTGGACCTGCTCCCTCGGGGGGCTCCTGAGCGACG	580
Mus	CTTCTCCCTCATCTCTAAAGCCAATGTGGATGTACTCCACGGAGGTCTCTGGAGCGCCA	593
Rattus	CTTCTCCCTCATCTCTAAAGCCAATGTAATGTACTCCACGGAGATCTCTGGAGCGCCA	575
Cricetulus	CTTCTCTCTCATCTCTAAAGCCAATATAGATTGCTCCACAGAGGTCTCTGGAGCGCCA	488
	***** * *. * *.***** *.** *.** ** ..**... *** ***,*.*.	

Homo GCGGCTGCTGCCTGCGGGCTC **TGGCCTGCTGG|GGTGTGCGGGGGT** TCTGCTGAGCGAGGC 707
 Pan GCGGCTGCTGCCTGCGGGCTC **TGGCCTGCTGG|GGTGTGCGGGGGT** TCTGCTGAGCGAGGC 701
 Nomascus GCGGCTGCTGCCTGCGGGCTC **TGGCCTGCGCG|GGCGTGC GGGGGT** TCTGCTGAGCGAGGC 700
 Macaca GAGGCTGCTGCCCGGGCTCTGACCTGCTGG|GGTGTGCGGGGGTCTCTGCTGAGCGAGGC 536
 Papio GAGGCTGCTGCCCGGGCTC **TGGCCTGCTGG|GGTGTGCAGGGGT** TCTGCTGAGCGAGGC 697
 Saimiri GCGACTGCTGCCCGGGGCC **TGGCCTGCGCG|GGCGTGC GGGGGG** TCTGCTGAGCGAGGC 640
 Mus GAGGCTGCTGATGGAGGCTCTGAAGTGCCAG|GGCGTGATGGAATTC AAGTGAGTGAGGC 653
 Rattus GAGGCTGCTGACCGGGCTCTGAAATGCCAG|GGTGTGATGGAATTC AAGTGAGTGAGAC 635
 Cricetulus AAGGCTGCTGCCT **AGAGCTCTGGAATGCCAG|GGCGTG TAT** AGCTTCAACTGAGTGAGAC 548
 ..*.***** .*. ** **.. *** .* ** *** .*. * *:.. **** ***. *

Homo TGATGTGCGGGCTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCCAGTC 767
 Pan TGATGTGCGGGCTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCCAGTC 761
 Nomascus TGATGTGACAGGCTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCTGAGTC 760
 Macaca TGATGTACGGGCTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCCAGTC 596
 Papio TGATGTACAGGCTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGTTTGTGGCCAGTC 757
 Saimiri CGACGTGCTGCTCTGGGAGGCCTGGCGTGCGACCTGCCC GGCGCTTTGTGGCCAGTC 700
 Mus AGATGTGCGGGCTCTCGAGGCCTGGCTGTGACCTGCCTGGGAAATTTGTGGCCAGATC 713
 Rattus GGATGCACGGGCTCTCGAGGCCTGGCTGTGACCTGCCTGGGAAATTCGTGGCCAAATC 695
 Cricetulus AGATGCGAAGGCTCTCGAGGCCTGGTCTGTGACCTGCCTGGAAATATGTGGTCAAAATC 608
 ** * ... ***** ***** ** ***** *. . *: **** ...**

Homo GGCCGAAGTGCTGTACCCCGGCTGGTGAGCTGCCCGGGACCCCTGGACCAGGACCAGCA 827
 Pan GGCCGAAGTGCTGTACCCCGGCTGGTGAGCTGCCTGGGACCCCTGGACCAGGACCAGCA 821
 Nomascus AGCAGAAGTGGTGCTACCCCGGCTGGCGCGCTGCCTGGGACCCCTGGACCAGGACCAGCA 820
 Macaca GGCAGAAGTGGTGCTACCCCGGCTGGTCCGCTGCTGGGACCCCTGGACCAGGACCAGCA 656
 Papio GGCAGAAGTGGTGCTACCCCGGCTGGTGCCTGCTGGGACCCCTGGACCAGGACCAGCA 817
 Saimiri GGCAGAGGCGGTGCTACCCCGGCTGGTGCCTGCTGGGACCCCTGGACCAGGAGCAGCA 760
 Mus TTCCGAAGTTCTCTCCCTGGCTGGCAGGATGCCAAGGACCCCTGGACCAGAGCCAGGA 773
 Rattus TTCGAAGTCTCTCTCCCTGGCTGGCAAGATGCGGAGGACCCCTGGACCAAGGCCAGGC 755
 Cricetulus TTCGAAGTCTCTCTCCCTGGCTGGCAGGCTGCCGAGGACCCCTGAATGAGGGCCAGGA 668
 * **.* * **.* ** ***** ..*** .*****.* *... *** .

Homo GGAGGC **AGCCAGGGCGGGCTCTGCAGGGCGGGG** GACCCCCCTACGG|CCCCCGTCGACATG 887
 Pan GGAGGC **AGCCAGGGCGGGCTCTGCAGGGCGGGG** GACCCCCCTACGG|CCCCCGTCGACATG 881
 Nomascus GGAGGC **AGCCAGGGCGGGCTCTGCAGGGCGGAG** GACCCCCCTACGG|CCCCCGTCGACGTG 880
 Macaca GGAAGC **AGCCAGGGCGGGCTCTGCAGAGAGGAG** GACCCCCCTACGG|CCCCCGTCAACGTG 716
 Papio GGA **AGCAGTCAGGGCAGCTCTGCAGAGAGGAG** GACCCCCCTACGG|CCCCCGTCAACGTG 877
 Saimiri GGAGGCGGGCC **GGGCAGCTCTGCAGGGCGGAG** GACCCCCCTACGG|CCCCCGTCGACGTG 820
 Mus AAAGGCAGTC **AGGGAGGTTCTGAGGAGTGGAA** GAACCAATATGG|CCCCCATCGAAGTG 833
 Rattus AAAGGCTGTC **AGGGAGGTTCTGAGGAGTGGAA** AGCCCCCTATGG|TCCCCATCGACGTG 815
 Cricetulus AG **AGGCAGTCAGGGAAGCTCTGAGGAGTGGAA** GATCCCCATACGG|CCCCCATCGAGGTG 728
 ..*.*** * *.***. * *****. *. ** ..** *****. ** ** ** ** ** ** ** ** ** ** ** ** *

Homo GTCTGTCTCCACGATGGACGCTCTGCGGGGCTGCTGCCCCGTGCTGGGCCAGCCATCAT 947
 Pan GTCTGTCTCCACGCTGGACGCTCTGCGGGGCTGCTGCCCCGTGCTGGGCCAGCCATCAT 941
 Nomascus GTCTGTCTCCACCTGGACGCTCTGACGGGCTGCTGCCCCGTGCTGGGCCAGCCATCAT 940

```

Macaca      GTCTATCTCCACCCTGGACGATCTGCAGAGCCTGTTGCCTGTGCTGGGCCAGCCCGTCAT 776
Papio       GTCTATCTCCACCCTGGACGCTCTGCAGAGCCTGTTGCCTGTGCTGGGCCAGCCCGTCAT 937
Saimiri     GTCAGCCTCCACCCTGGACGCTCTGCGGGGCCTGCTCCCCGTGCTGGGCCAGCCCGTCGT 880
Mus         GTCAGTCTCCACCCTGGATGCCCTGCAGAGCTTGGTAGCAGTGTGGATGAGTCCATCGT 893
Rattus      GTCAGTCTCCACCCTGGATGCCCTGCAGGGTTTGGCTGGTAGTGTGGATGAGTCCATTGT 875
Cricetulus  GTCAGTCTCCACCCTGAATGCTCTTCAGGGCCTGCTGACAGTGTGGATGAGTCCATCAT 788
***:. ***** .**.* *. ** *.**.* ** * *** ***. .* **.* .*

```

Nella sezione mostrata, relativa agli esoni dal 6 all'8 (che per *Homo sapiens* corrispondono agli esoni dal 7 al 9) sono evidenziati in giallo i tre siti a maggiore energia di legame identificati inizialmente per *Homo sapiens*; in magenta invece è messo in evidenza il blocco di siti legame predetti per le singole specie, e che si è rilevato filogeneticamente conservato in modo quasi perfetto.

Un secondo seed-match è risultato particolarmente conservato tra le diverse specie, ed è stato messo in evidenza con il colore rosso nell'allineamento a blocchi.

F.0.2 Blocco2

```

Esone              11    12
Odobenus           GTCAGCAAAGGCACAGACAATGGATGCTCAG | GTGGCTGCCCTGATTGC CCG 581
Mustela            GTCAGCAAAGGCAGAGAAGAGGACGCTCAG | GTGGCTGCCCTGATCGC CCG 584
Canis              GTCAGCAAAGGCACAAAAGTGGATGCTCAG | GTGGCTGCCCTGATTGC CTG 584
Felis catus        GTCAGCAAAGGCACGAAATGGATGCTCAG | GTGGCTGCCCTGATTGC CCG 584
Equus              ATCAGCGAGGGGCGCAGAATGGATGCTCAG | GTGGCTGCTCTGATTGCCCG 584
Bos                GTCAGCCAAGGGCGTGGTGTGGATGCTCAG | GTGGCTGCCCTGATTGCCCG 584
Sus                GTCAGCAAAGGGCGGGGGATGGATGCACAG | GTGGCCGCCCTGATTGC CCG 584
Homo               GTCAACAAAGGGCACGAAATGAGTCCTCAG | GTGGCCACCCTGATCGACCG
*** * * * *      *   * *** ***** ***** * * *

Odobenus           CTATGTGGCGGGAGGGGGCCA GCTGGACAAGGCCTCTTTGGACACTGG 631
Mustela            CTATGTGGCGGGAGGGGGCAA GCTGGACCAGGCCACCCTGGACATGCTGG 634
Canis              CTACGTGGCAGGAGGGGG CCAACTGGACAAGGCCACCCTGGACGTACTGG 634
Felis catus        CTATCTGGTGGGAGGGGGC CAGCTGGACAAGGCCACCCTGGACACTGG 634
Equus              CTACC TGGTGGGAGGGGGCCAGATGGACA AGGCCACCCTGGACACTGG 634
Bos                CTATGT AGGGGGAGGAGGCCAGCTGGATA AGGCCGCCCTGGACACCCTGG 634
Sus                CTATGTAGGGGAGGGGG CCAACTAGACGGCGCCACCCTGGACATCCTAG 634
Homo               CTTTGTGAAGGGAAGGGGCCAGCTAGACAAAGACACCCTAGACACCCTGA 1300
**   *   *** * * * * * * *   * * * * * * *

Odobenus           CCACCTTCGGCCCCACGTACCTGTGTTTCCTCAGTCCTGAGCAGCTGAGC 681
Mustela            CCACCTCAGGCCCCGCATACCTGTGTTTCCTCAGTCCTGAGCAGCTGCGC 684

```

Esone	12	13
Odobenus	TCCGTGCAACCCAGCGTCCTTTG	GGCAGCCAGGCCCCAGGACCTGGAGGC 731
Mustela	TCCGTGCGACACAGCGTCCTTTG	GGCAGCCGGGCCCCAGGACCTGGATGC 734
Canis	TCTGTGCGACACAGTGTATTTG	GGTGGCCAGGGCCCAGGACCTGGATGC 734
Felis catus	TCAGTGCACCACAGCGTTGTTTG	GGCAGCCCGGCCCCCGGATCTGGATGC 734
Equus	TCCATGCCACACAACGTTGTTTG	GGCGGCCGGGCCCCAGGACCTGGACAC 734
Bos	TCCGTACAGCTCAGCGTGCTTTG	GATGACCACACCCAGGACCTGGACGC 734
Sus	TCTATACAGCTCAGCGTTGTTTG	GGTGACCAGGCCCTTGACCTGGACAC 734
Homo	TCCGTGCCCCCAGCAGCATCTG	GGCGGTGACGGCCCTAGGACCTGGACAC 1400

Species	Sequence	Position	GC Content (%)
Esone	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	14	15
Odobenus	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	931	
Mustela	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	934	
Canis	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	934	
Felis catus	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	934	
Equus	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC	934	
Bos	TATAGACGGCGGGCATTCAAGAACTGCAGACAGAGCGGTGCTG TCGC	934	
Sus	CCTGGACATCGCCACATTTAAGAACTGCAGACGGAGGTCTTTCGG AAGC	934	
Homo	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGTGCTG CCGT	1600	

Odobenus	TGACCATTGCCGAGGTGCAAAATCTTCTGGGTCCAAACCTAGT	GGGC	978
Mustela	TGACCATTGCCGAGGTGCAAAACCTTCTGGGTCCAAACCTGGT	GGGC	981
Canis	TGACTGTTGCCGAGGTGCAAAACCTTTTGGGTGCAAAACCTGGC	GGGC	981
Felis catus	TGACCGTTGCCGAGGTCCGAAACCTTCTGGGTCCGAACTGGT	GGGC	981
Equus	TGACCATTGCCGAGGTGCAAAAACCTTCTGGGTCCGAACTGGC	GGGC	981
Bos	TAACCGTGCCGAGGTGCAGAAACTTCTGGGCCCAAACCTGGTAGG	C	981
Sus	TGACCATCGCTGAGGTGCAGAAACTTCTGGGTCCACACTGGCGGGC		981
Homo	TGACTGTGGCTGAGGTGCAGAAACTTCTGGGACCCACAGTGGGA	GGGC	1647
	* * * * *		

Odobenus	CTGAAGGCAGAGCAGGAGAGCAG	CCCCGTGCGGGACTGGATCTTACGGCAGCG	1031
Mustela	CTGAAGGCGGAGCAGGAGAGGAG	CCCCGTGCGGGACTGGATAGTCCGGCAGCA	1034
Canis	CTGAAGGCGGAGCAGGGGAA	CAGCCCCGTGCGGGACTGGATCTTACGGCAGCG	1034
Felis catus	CTGAAGGCGGCGCGGGAGA	GCAGCCCCGTGCGGGATTGGATCTCACGGCAGCG	1034
Equus	CTGAAGGCTGAGGAGGGCAA	CAGCCCCGTGCGGGACTGGATCTCCGGCAGCC	1034
Bos	CTGAAGGCTGAGACAGGGAA	CATGTTCTGCGGGACTGGATATCCCGGCAGTC	1034
Sus	CTGAAGGCGGAGGAGGAGAA	CAGCCCTGTGCGCAGTGGATCTCCCGGCAGCC	1034
Homo	CTGAAGGCGGAGGAGCGGCACGCCGGTGCGGGACTGGATCCTACGGCAGCG		1700


```

***** * *
***** ** *****
*****

Odobenus      GCAGGATGACCTGGACAGTCTGGGCCTGGGGCTCCACGGTGGCATCCCCA 1081
Mustela       GCAGGATGACCTGGACAGTCTGGGCCTGGGGCTGCACGGTGGCCTCCCCA 1084
Canis         GCAGGATGACCTGGACCGTCTGGGCCTGGGGCTCCGCGGTGGCATCCCCA 1084
Felis catus   GCAGGAGGACCTGGACAGTCTGGGCCTGGGGCTCCGAGGCGGCATCCCCA 1084
Equus        GCAGGACGACCTGAACAGTCTGGGTCTGGGTCTCCGCGCGGCATCCCCA 1084
Bos          GCAGGAGGACCTGGACAGGCTGGGGCTGGGGCTCGTCGCGCGGCATCCCCA 1084
Sus          GCAGGACGACCTGGACCGGCTGGGGCTGGGGCTCCAGGCGGCATCCCCA 1084
Homo         GCAGGACGACCTGGACACGCTGGGGCTGGGGCTACAGGCGGCATCCCCA 1750
***** ***** ** ***** ** ** ** *****

Esone
Odobenus      15 16
               ATGGCTACCTGGTCATGGACCTTAGCTTCCGAG|AGGCCCTCTCAGGGGGC
Mustela       ATGGCTACCTGGTCATGGACCTCAGGCTCCGAG|AGGCCCTCTCAGGGGGC 1134
Canis         ACGGCTATCTGGTCATGGACCTCAGCTTCCGAG|AGGCCCTCTCAGGGGGC 1134
Felis catus   ATGGCTACCTGGTCGTGGACCTCAGCTTCCGAG|AGGCCCTCTCAGGGGGC 1134
Equus        ACGGCTACCTGGTCCTGGACCTCAGCTTCCGAG|AGGCCCTCTCAGGGGGC 1134
Bos          ACGGCTACCTGGTCCTGGACCTCCGCGGACGAG|AGGCCTCCTCGGGAGGC 1134
Sus          ACGGCTACCTGGTCCTGGACATCAACATCCGAG|AGGCCCTCTCGGGAGGC 1134
Homo         ACGGCTACCTGGTCCTAGACCTCAGCATGCAAG|AGGCCCTCTCGGGGACG 1800
* ***** * * * * * * * * * * * * * * * *

```


Allineamenti proteici e del cDNA per la mesotelina

G.1 Allineamenti proteici per la sequenza completa

Blocco 1

Splicing phase		1		0	
Exon number		1	2	2	3
Homo_sapiens	MALPTARPLLGSCGTPALG-SLLFLLFSL GWVQPSRTLAGETGQ EAAPLD	49			
Pan_troglodytes	MALPTARPLLGSCGTPALG-SLLFLLFSF GWVQPSRALAGETGQ EAAPLD	49			
Nomascus_leucogenys	MALLTARSLLGSCGTPALS-SLLLLLFSL GWVQPSRALAGETGQ EAAPLD	49			
Macaca_mulatta	MALPMARPLSGSCGTPALG-SLLFLLFSL GWVQPSRVLAGETRQ EAAPLD	48			
Papio_anubis	MALPTARPLLGSCGTSALG-SLLFLLFSL GWVQPSRALAGETRQ EAAPLD	49			
Saimiri_boliviensis	MALPTARPLWRSCGSPALG-RLPFLLLSL GWVQPWKAWAGETGQ EAAPLD	49			
Mus_musculus	MALPTARPLLGSCGSPICSRFLLLLLSL GWIPRLQTQTTKTSQ EAT-LL	49			
Rattus_norvegicus	MALPTAQPLLGSCGSPICSRFLLLLLSL GWLPLLQTQTTKTSQ EAA-LL	49			
Cricetulus_griseus	MALPTARPLLGSCGSPICCSLLLLLSL GWMPVLQLQATKTDQ EAVHLC	50			
	*** *. * ***. : :*: *: ** : : . * :*. *				

Splicing phase		2	
Exon number		3	4
Homo_sapiens	GVLANPPNISS LSPRQLLGFPFCAEVSGLSTERVRELAVALAQKNVKLSTE	99	
Pan_paniscus	GVLANPPNISS LSPRQLLGFPFCAEVSGLSTERVRELAVALAQKNVKLSAE	99	
Nomascus_leucogenys	GVLNAPNIAS LSPRQLLGFRCAEVSGLSTERVQELAVALGQKNVKLSAE	99	
Macaca_mulatta	GILTNPDIAS LSPRQLLGFTCEVSGLSTELVQELAVALGQKNVKLSAE	98	
Papio_anubis	GVLNAPDIAS LSPRQLLGFTCEVSGLSTELVQELAVALGQKNVKLSAE	99	
Saimiri_boliviensis	RVLNARNIAS LSPGGLLGFTCAEVSALESVWELAVSLGQKNVTLAE	99	
Mus_musculus	HAVNGADFAS LPTGLFLGLTCEEVSDLSMEQAKGLAMAVRQKNITLRGH	99	
Rattus_norvegicus	HAVTGTVDFAAS LPTGLFLGLTCEEVSGLSMGAHELAMAVRQKNIVLQVH	99	
Cricetulus_griseus	DELIGNPDFAS LPAGLFLGLTCEEVSGLSMAHAQALAMAVRQKNITLRVN	100	
	: . :*: *.. :*: * *** * . **: :*: * .		

Splicing phase	0	2	
Exon number	4 5	5 6	
Homo_sapiens	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTRFFFSRITKA 147		
Pan_paniscus	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTRFFFSRVTKA 147		
Nomascus_leucogenys	Q LRCLAHRLSE--PPKDLDALPLDLLLFLN PDAFSGPRACTRFFFSRVTRA 147		
Macaca_mulatta	Q LRCLAHQLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTHFFFSRVAKA 146		
Papio_anubis	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDTFSGPQACTHFFFSRVAKA 147		
Saimiri_boliviensis	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PAAFSGSQACTRFFFSRVSKA 147		
Mus_musculus	Q LRCLARRLPKHLTDEELNALPLDLLLFLN PAMFPGQQACAHHFSLISKA 149		
Rattus_norvegicus	Q LRCLARRLPKHLTNEELDALPLDLLLFLN PAMFPGQQACAHHFSLISKA 149		
Cricetulus_griseus	Q LRCLARRLPKYLTNEELDALPLDLLLFFN PAMFRGQQACVHHFSLISKA 150		
	* *****:*. . . :*:*** *****:* * * * :**.:*** :*:*		

Splicing phase	0	
Exon number	6 7	
Homo_sapiens	NVDLLPRGAPERQRLIPAALACW GVRGSLLEADVRALGGLACDLPGRFV 197	
Pan_paniscus	NVDLLPRGAPERQRLIPAALACW GVRGSLLEADVRALGGLACDLPGRFV 197	
Nomascus_leucogenys	NVDLLPREAPERQRLIAAALACR GVRGSLLEADVQALGGLACDLPGRFV 197	
Macaca_mulatta	NVDLLPRGAPERQRLIPGALTWC GVRGSLLEADVRALGGLACDLPGRFV 196	
Papio_anubis	NVDLLPRGAPERQRLIPAALACW GVQGSLLSEADVQALGGLACDLPGRFV 197	
Saimiri_boliviensis	NVDLLPRGAPERRRLLPAALACR GVRGALLSEADVRALGGLACDLPGRFV 197	
Mus_musculus	NVDVLPRRSLERQRLIMEALKCQ GVYGFQVSEADVRALGGLACDLPKGFV 199	
Rattus_norvegicus	NVNVLPRRSLEQRLTGALKCQ GVYGFQVSETDARALGGLACDLPGEFV 199	
Cricetulus_griseus	NIDLLPQRSLEQRLIPRALECQ GVYSFQLSETDAKALGGLVCDLPKGYV 200	
	:::: : **:*** ** * ** . :*:*.:*****.*****.:*	

Splicing phase	2	
Exon number	7 8	
Homo_sapiens	AESAEVLLPRLVSCPGPLDQDQQAARAALQGGGPPYG PPSTWSVSTMDA 247	
Pan_paniscus	AESAEVLLPRLVSCGLPLDQDQQAARAALQGGGPPYG PPSTWSVSTLDA 247	
Nomascus_leucogenys	AESAEVLLPRLARCLGPLDQDQQAARAALQGGGPPYG PPSTWSVSTLDA 247	
Macaca_mulatta	AESAEVLLPRLVRCLGPLDQDQQAARAALQGGGPPYG PPSTWSISTLDD 246	
Papio_anubis	AESAEVLLPRLVRCLGPLDQDQQAARAALQGGGPPYG PPSTWSISTLDA 247	
Saimiri_boliviensis	AESAEVLLPRLVHCLGPLDQDQQAARAALQGGGPPYG PPSTWSASTLDA 247	
Mus_musculus	ARSSEVLLPWLACGQGPLDQSQEKAVREVLRSGRTPYG PPSKWSVSTLDA 249	
Rattus_norvegicus	AKSSEVLLPWLARCGLDQGQAKAVREVLRSGRAPYG PPSTWSVSTLDA 249	
Cricetulus_griseus	VKSSEVLLPWLACRGLNEGQEEAVREALRSGRSPYG PPSRWSVSTLNA 250	
	..*:*.:** *. * ***:: * :*. . *: * . ** *** ** ***::	

Splicing phase	0	
Exon number	8 9	
Homo_sapiens	LRGLLPVLGQPIIRSIPQ GIVAARQRSSRDPSWRQPETILRPRFRREV 297	
Pan_paniscus	LRGLLPVLGQPIIRSIPQ GIVAARQRSSRDPSWQQPERTILWPRFRRDV 297	
Nomascus_leucogenys	LQGLLPVLGRPIIRSIPQ GMVARVAATLPWDPSWRQPEQTXLRPRFRRDV 297	
Macaca_mulatta	LQSLLPVLGQPVISARPQ GILAARQRSSRDPSWQQPEQTVLRLRFRRDV 296	
Papio_anubis	LQSLLPVLGQPVHSIPQ GILAARQRSSRDPSWRQPEQTILRPRFRRDV 297	
Saimiri_boliviensis	LRGLLPVLGQPVVRSIPQ GVVAARKRTRSDPSWRQPERTAALRPLRRDM 297	
Mus_musculus	LQSLVAVLDESIVQSIK DVKAEWLQHSIRDPSRLGSKLTVIHPFRRDA 299	
Rattus_norvegicus	LQGLLVVLDSEIVHSIPK DVITEWLQGISREPSRLGSKWTVTHPRFRDRT 299	
Cricetulus_griseus	LQGLLVVLDSEIHSIPK DVKAEWLQHSIRGSPSWRGIKLTAMIPRFRDRT 300	

		.:: **...: : *: .: : . ** : * *:***:
Splicing phase	1	
Exon number	9 10	
Homo_sapiens	E KTACPSGKKAREIDESLIFYKKWELEACVDAALLATQMDRVNAIPFTYE	347
Pan_paniscus	E KTACPSGKKAREIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE	47
Nomascus_leucogenys	E RTACPPGKEAHEINENLIFYKKWELEACVDAALLAQMDHVNAIPFTYE	47
Macaca_mulatta	E RTTCPPEKEVHEIDESLIFYKKRELEACVDPALAAQMDRVDAIPFTYE	346
Papio_anubis	E RTTCPPEKEVHEIDESLIFYKKRELEACVDAALLAQMDRVNAIPFTYE	347
Saimiri_boliviensis	E KKACPPGKEVLEIDENLIFYKNWELEACVNASLLATQIDRVDAIPFTYE	347
Mus_musculus	E QKACPPGKEPYKVDDELIFYQNWELEACVDGTLARQMDLVNEIPFTYE	349
Rattus_norvegicus	E QKACPPGKEPNVDENLIFYQNWELEACVDGTLLAGQMDLVNEIPFTYE	349
Cricetulus_griseus	E RKACPPGMEPHKVDENLIFYQNWELEACVDGAMLATQMDLVNEIPFTYE	350
	* :.:** . : :*:*****: *****: :** *: * :*****	
Splicing phase	0	
Exon number	10 11	
Homo_sapiens	QLDVLKHKLDE LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLETLKA	397
Pan_paniscus	QLDVLKRKLDE LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLETLKA	397
Nomascus_leucogenys	QLDVLKHKLDE LYPQGYPESVIQHLDHLFLKMSPEDIRKWNVTSLETLKA	397
Macaca_mulatta	QLDVLKHKLDE LYPQGYPESVIRHLGHLFLKMSPEDIRKWNVTSLETLKA	396
Papio_anubis	QLDVLKHKLDE LYPQGYPESVIRHLGHLFLKMSPEDIRKWNVTSLETLKA	397
Saimiri_boliviensis	QLGILKHKLDE LYPQGYPEPVIQHLGQFFLKMSPEDIRKWNVTSLETLKA	397
Mus_musculus	QLSIFKHKLDK TYPQGYPESLIQQLGHFFRYVSPEDIHQWNVTSPTVKT	399
Rattus_norvegicus	QLSIFKHKLDK TYPQGYPESLIQQLGHFFRYVSPEDIRQWNVTSPTVNT	399
Cricetulus_griseus	QLDIFKHKLDE TYPQGYPESLIQQLGHFFRYISPEDIHRWNVSSPTVKT	400
	.:*:*: *****.:*:*. :* :*****:***:* :*:::	
Splicing phase	0	
Exon number	11 12	
Homo_sapiens	LLEVNGHEMSPQ VATLIDRFVKGQQLDKDITLDTLTAFFPYGLCSLSPE	447
Pan_paniscus	LLEVNGHEMSPQ VATLIDRFVKGQQLDKDITLDTLTAFFPYGLCSLSPE	447
Nomascus_leucogenys	LLEVSKGHKMSAQ VATVIDRFVMGRQQLDKDITLDTLTAFFPYGLCSLSPE	447
Macaca_mulatta	LLKVSKGHEMSAQ VATLIDRVVVGQQLDKDITVDTLTAFCPGCLCSLSPE	446
Papio_anubis	LLKVSKGHEMSAQ VATLIDRVVVGQQLDKDITVDTLTAFCPGCLCSLSPE	447
Saimiri_boliviensis	LLKVSKGHEMSVQ VATLIDHFLMGRRQLDKDITLDTLTAFFPYGLCSLSPE	447
Mus_musculus	LLKVSKGQKMNAQ AIALVACYLRGGGQLDEDIVKALDNIPLSYLCDFSPQ	449
Rattus_norvegicus	LLKVSKGQKMDAQ VIALVACYLRGGGQLDEDIVKALDNIPLSYLCDFSPQ	449
Cricetulus_griseus	LLKVTKGQKMDAQ VTALVACYLRGGGHLQDIVKAMDDIPLTYLCDFKPKQ	450
	.:.:*. * . : : : * :***: * : : : * :***:*	
Splicing phase	2	
Exon number	12 13	
Homo_sapiens	ELSSVPPSSIW AVRPQDLDTCDPRQLDVLYPKARLAFQNMNGSEYFVKIQ	497
Pan_paniscus	ELSSVPPSSIW AVRPQDLDTCDPRQLDVLYPKARLAFQNMNGSEYFVKIQ	497
Nomascus_leucogenys	QLSSVPPSSIW AIRPQDLDTCGPRQLDVLYPKARLAFQNMNGSEYFVKIQ	497
Macaca_mulatta	RLSSVPPSVIG AVRPQDLDTCGPRQLDVLYPKARLAFQNMMSGSEYFVKIR	496
Papio_anubis	RLSSMPPSIIG AVRPQDLDTCGLRQLDVLYPKARLAFQNMMSGSEYFVKIR	497
Saimiri_boliviensis	QLGSVPPSLIW AVRPQDLDTCGPRQLEVLYPKARLAFQNVNGSEYFSKIR	497
Mus_musculus	DLHSVPSVMW LVGPQDLDKCSQRHLGLLYQKACSAFQNVSGLEYFEKIK	499

```
Rattus_norvegicus      DLHAIPSSVMV|LVGLHDLKCSQRHLGILYQKACSAFQNVSGLEYFEKIR 499
Cricetulus_griseus    NLHSIPSSFMV|LVKPDNLNCSQRHLDILYQKACLAFAQNSSRQYEFKIR 500
*  ::*:  :  :  ::*:  *  *:  ***  **  *****  .  ****  **:
```

Splicing phase	1	0
Exon number	13 14	14 15
Homo_sapiens	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL	PLTVAEVQKLLGPHV 547
Pan_paniscus	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL	PLTVAEVQKLLGPHV 547
Nomascus_leucogenys	SFL GGAPTEDLKALSQQNASMDLATFMKLRTDAVL	PLTVAEVQKLLGPHV 547
Macaca_mulatta	PFL GGAPTEDEVKALSQQNVSMDLATFMKLREAVL	PLTVAEVQKLLGPHV 546
Papio_anubis	PFL GGAPTEDLKALSQQNVSMDLATFMKLREAVL	PLTVAEVQKLLGPHV 547
Saimiri_boliviensis	SFL GGAPAECLKLSQRNVSMDLATFMKLRTAEVL	PLTVAEVQKLLGPHV 547
Mus_musculus	TFL GGASVKDLRLALSQHNVSMDIATFKRLQVDSL	GLSVAEVQKLLGPNI 549
Rattus_norvegicus	TFL GGASREDLRALSQHNVSMDIATFKKLQVDAL	GLSVAEVQKLLGPHI 549
Cricetulus_griseus	TFL GGASVEDLRALSQQNLMSDLATFKKLQVDAL	GLTVAEVQKLLGPHI 550
	..** ***. :::: ***:* ***:*** :*: :::: *.*****:::	

Splicing phase		1
Exon number		15 16
Homo_sapiens	EGLKAEERHRPVRDWILRQRQDDLDTLGLGLQGGIPNGYLVLDLSMQ EAL	597
Pan_paniscus	EGLKAEERHPPVRDWILRQRQDDLDTLGLGLQGGIPNGYLVLDLSVR EAL	597
Nomascus_leucogenys	EGLKAEERHSPVRDWILGQRQDDLDAALGLGLQGGIPNGYLVLDVSVR EAL	597
Macaca_mulatta	EGLKVEEQHSPVRDWILKQRQDDLDTLGLGLQGGIPNGYLILDLFSVR EAL	596
Papio_anubis	EGLKVEEQHSPVRDWILKQRQDDLDTLGLGLQGGIPNGYLILDLFSVR EAL	597
Saimiri_boliviensis	EGLKAEERHSPVRDWILRQRQDDLDTLGLGLQGGIPNGYLILDRGVR EAL	597
Mus_musculus	VDLKTEEDKSPVRDWLFRHQKQDLRLGLGLQGGIPNGYLVLDNFVR EAF	599
Rattus_norvegicus	GLDKTEEDKSPVRDWLFRQQQKDLDSLGLGLQGGIPNGYLILDFNVR EAF	599
Cricetulus_griseus	ADLKTEEDKSPVRDWLFRQREQEDLRLGLGLQGGIPNGYLILDLNVR EAF	600

.**.* : ***** : *.**.* *****.*** : : ** :

Homo_sapiens	SGTPCLLGPGPVLTVLALLLASTLA-	622
Pan_paniscus	SGTPCLLGTPVLTVLALLLASTLA-	622
Nomascus_leucogenys	SGTPCLLGPAVLTVLALLLASTLA-	622
Macaca_mulatta	SGTPCLLGPGPVLTVLALLLASTLA-	621
Papio_anubis	SGTPCLLGPGPVLTVLALLLASTLA-	622
Saimiri_boliviensis	SGTPCLLGPGPVLTVLALLAPTAC	623
Mus_musculus	SSRASLLGPGFVLWIPALLPALRLS	625
Rattus_norvegicus	SSGAPLLGPGFVFAWIPALLSALRLS	625
Cricetulus_griseus	SSGAPLLGPGFVIAWIPALLSAFILS	626
	*.	

BLOCCO 2

```

Splicing phase                      1                      0
Odobenus_rosmarus  MSLQTARPPPLGSCGTSAHGNLLLLLLSL|GWVFPsRAQAADSRL|----- 41
Mustela_putorius   MAMRTAQPAWGSWGTsAGGSLLLLLLSL|GWAFPSRAQAADLRL|----- 41
Canis_lupus        MALQTAQPPLGSRWTSTRGSLLLLLLSL|GWVLPsRAQAADSRL|----- 41
Felis_catus        MALQAALPPLGSRGTPTHCSLLLLLLSL|GWVLPsRVRAADSRL|----- 41
Bos_taurus         MALQTAPPRLGSWGTsPAHGSLLLLLLCL|GWVLPsRAQAADVGL|----- 41
Sus_scrofa         MALQTARPHLGSRGTRTPRSLLLLLLSL|GWVPPsRAQAADTGQ|----- 41
Equus_caballus     MALQTAQRPLGSCGTPTHGSLLLLLLSL|GWLPPPGAQAADTGL|----- 41
Homo_sapiens       MALPTARPLLGSCTPALGSLLFLLFSL|GWVQPSRTLAGETGQ|EAPLDGVLANPPNISS 60
*:::*  ** * : .**:**.* ** *. . *.:

```

```

Splicing phase                      2                      0
Odobenus_rosmarus  -----
Mustela_putorius   -----
Canis_lupus        -----
Felis_catus        -----
Bos_taurus         -----
Sus_scrofa         -----
Equus_caballus     -----
Homo_sapiens       |LSPRQLLGFPsCAEVsGLSTERVRELAVALAQKNVKLSTEQ|LRCLAHRLSEPPEDLDALPL 120

```

```

Splicing phase                      2                      0
Odobenus_rosmarus  -----
Mustela_putorius   -----
Canis_lupus        -----
Felis_catus        -----
Bos_taurus         -----
Sus_scrofa         -----
Equus_caballus     -----
Homo_sapiens       DLLLFLN|PDAFSGPQACTRFFsRITKANVDLLPRGAPERQRLLPALACW|GVRGSLLSEA 180

```

```

Splicing phase                      2
Odobenus_rosmarus  -----
Mustela_putorius   -----
Canis_lupus        -----
Felis_catus        -----
Bos_taurus         -----
Sus_scrofa         -----
Equus_caballus     -----
Homo_sapiens       DVRALGGLACDLPGRFVAESAEVLLPRLVSCPGPLDQDQQAARAALQGGGPPYG|PPSTW 240

```

```

Splicing phase                      0                      1 (2)
Odobenus_rosmarus  -----|GVMPWLQGTsRDPsWQPELTvVLLRRRDTE|-K 77
Mustela_putorius   -----|GVMPWLQGTsRVLsWWRPELSVVLsRRDRDID|SK 78
Canis_lupus        -----|GVMPWLQGTsRDLTWQPELTvILsRRDRDID|KK 78

```

```

Felis_catus      -----|GVMPWLQGTSRDLWQRPVLTLLPRDRRDTE|KK 78
Bos_taurus      -----|ALLSSSVQRTSRDLWQQPELTVLLPRARRGTE|KK 78
Sus_scrofa      -----|GVLIPWLQRTSGDLWQPPEPTAIRPRPQGGTE|KR 78
Equus_caballus  -----|GFVTPWLQRPTRDPWSWQQPELTVVLLQRAQRSTK|S(|)K
Homo_sapiens    SVSTMDALRGLLPVLGQPIIRSIPQ|GIVAAWRQSSRDPSWRQPVERTILRPRFRREVE|KT 300

```

```

..: . * .: :* ** : : * : ..

```

```

Splicing phase
Odobenus_rosmarus VCPSEKKAHVIDENLVFYEEWELEACVDGALLAAQMDRVNMIPFTYQQLDVFVKRKLDE|FY 137
Mustela_putorius VCPPSRKAHVIDENLVFYEEWELEACVDGALVAAQMDRVNMVPFTYQQLDVFVKRKLDE|FY 138
Canis_lupus      VCPPGQKAHVVDENLVFYEEWELEACVDGALLAAQMDRVNMVPFTYQQLDVFVKRKLDE|LY 138
Felis_catus      ACPPERRARVVDENLVFYEPWELEACVDGALLAAQMDRVNMVPFTYQQLGVFVKRKLDE|LY 138
Bos_taurus      ACPPGREAQVVDENLVFYEEWELEACVDGALLATQMRQVNLIPFTYQQLHIFVKRKLDE|FY 138
Sus_scrofa      DCPPEGKRAQVVDENLVFYEDWELEACVDGALLADQMDQVNLIPFTYQQLHVFVKRKLDE|LY 138
Equus_caballus  ACPEPEKATVVDENLVFYEEWELEACVDAALLDAQMDRVNAIPFTHQQLEIFRCKLNQ|FY 138
Homo_sapiens    ACPGKKAKEIDESLIFYKKWELEACVDAALLATQMDRVNAIPFTYQQLDVLKHKLDE|LY 360

```

```

** . : * : ** * : ** : ***** : ** : * : ** : *** : ** : : : *** : *

```

```

Splicing phase
Odobenus_rosmarus PQGYPESLIQHLRYFFLELTPADIKWNVTSLETVKSLQVSKGQTMDAQ|VAALIARYVA 197
Mustela_putorius PQGYPESLTQNLGYFFLELSPADIKWNVTSLETVKSLQVSKGQKRDQ|VAALIARYVA 198
Canis_lupus      PQGYPESLIQHLGYFFFEITPEDIHKWNVTSLETVKSLKVSQGHKVDQ|VAALIACYVA 198
Felis_catus      PQGYPESLVRHLGYFFLELTPEDIHKWNVMSLETVKSLLEVSKGHEMDQ|VAALIARYLV 198
Bos_taurus      PQGYPQSLIERLSYFFLYVTPEDIHKWNVTSLETVKSLKVSQGRGVDAQ|VAALIARYVG 198
Sus_scrofa      PQGYPTSLVQRLRYFFLSVTPEDICKWNVTSLETVKSLXVSKGRGMDQ|VAALIARYVG 198
Equus_caballus  PQGYPESLIRRLRYFFYAVTPEDIYKWNVTSPYIVKYLKISEGRMDQ|VAVLIARYLV 198
Homo_sapiens    PQGYPESVIQHGLGYLFLKMSPEDIRKWNVTSLTLKALLEVNKGHEMSPQ|VATLIDRFVK 420

```

```

**** * : .. * : * : ** * * * * * : : * * * : : : : .. * * * * : :

```

```

Splicing phase
Odobenus_rosmarus GGGQLDKASLDTLATFGPTYLCFLSPEQLSSVQPSVLW|AARPQDLEACRPLQMDVLYPRA 257
Mustela_putorius GGGQLDQATLDMLATFRPAYLCFLSLEQLRSVQHSVLW|AAGPQDLDACHPQMAVLYPRA 258
Canis_lupus      GGGQLDKATLDVLTATFSPTYLCVLSPEQLGVSQHSVIW|VARAQDLDACRPPQMDVLYPKA 258
Felis_catus      GGGQLDKATLDTLAHFPTYLCFLSPEQLESVHHSVWV|AARPDLDACRPAQMDVLYRKA 258
Bos_taurus      GGGQLDKAALDTLAIFRPAYLCFLRPEQLDSVQLSVLW|MTTPQDLDACSPQMAVLYHKA 258
Sus_scrofa      GGGQLDGATLDILATFCPTYLCFLSPEQLSSIQLSVVR|VTRPLDLDTCSPLQVAVLYLKA 258
Equus_caballus  GGGQMDKATLDTLAIFRPTYLCFLSPEQLDSMPHNWVW|AAGPQDLDTCRPPQMDALYSKA 258
Homo_sapiens    GRGQLDKDTLDTLATFYPGYLCSLSPEELSSVPPSSIW|AVRPQDLDTCDPRQLDVLYPKA 480

```

```

* * : : * : ** * : * * * * * * : * * : . : . . * : * * * : * * : * : *

```

```

Splicing phase
Odobenus_rosmarus RAAFQNINGSEYFVKIKPYL|AGAPTEDLRALSRRKVNMDMATFKTLQTEAVL|PLTIAEVQ 317
Mustela_putorius RTAFQTVSGPEYFSRIKPFL|GGASTEDLRALSSQKVNMDVATFQKLQTEAVL|PLTIAEVQ 318
Canis_lupus      RTAFQNLGSEYFERIKPYL|GGARTEDLRALSQQNVMSMDTATFKTLQREAVL|PLTVAEVQ 318
Felis_catus      RLAFQNMMSGSEYFEKMKLYL|GGAPTEDLRALSQRNISMDLAAFRTLRPEAVV|PLTVAEVR 318
Bos_taurus      HMAFQNVSGSEYFTRIKPFL|GGASTEDLRVFTSQNISIDAAAFKKLTTEAVL|SLTVAEVQ 318
Sus_scrofa      RAAFQNMMSGSEYLARIKPYL|GGAPTEDLRALSQRNISLDIATFKKLQTEVLR|KLTIAEVQ 318
Equus_caballus  RIAFQNMMSGSEYFARIKPYL|GGAPTEDLQALSQRNIHMDMATFKKLRTTEAVL|PLTIAEVQ 318
Homo_sapiens    RLAFQNMMSGSEYFVKIQSFL|GGAPTEDLKALSQQNVMSMDLATFMKLRTDAVL|PLTVAEVQ 540

```

```

: * : : * : ** * : : : * : * * * : : : : * : * * : : : * : * : * :

```


Splicing phase		1
Odobenus_rosmarus	NLLGPNLVGLKAEQESSPVRDWILRQRQDDLDLSGLGLHGGIPNGYLVMDLSFR EALSGG	377
Mustela_putorius	NLLGPNLVGLKAEQERSPVRDWIVRQQDDLDLSGLGLHGGIPNGYLVMDLRLR EALSGG	378
Canis_lupus	NLLGANLAGLKAEQGNPVRDWILRQRQDDLDRLGLGLRGGIPNGYLVMDLSFR EALSGG	378
Felis_catus	NLLGPNLVGLKAAESSPVRDWISRQRQEDLDLSGLGLRGGIPNGYLVVDLSFR EALSGG	378
Bos_taurus	KLLGPNLVGLKAETGNMFLRDWISRQSQEDLDRLGLGLVGGVPNGYLVLDLRGR EASSGG	378
Sus_scrofa	KLLGPHLAGLKAEENSPVRDWISRQPQDDLDRLGLGLGGIPNGYLVLDINIR EAPWGG	378
Equus_caballus	KLLGPNLAGLKAEENSPVRDWILRQPQDDLNSLGLGLRGGIPNGYLVLDLSFR EALSGG	378
Homo_sapiens	KLLGPHVEGLKAEERHRPVRDWILRQRQDDLDLGLGLQGGIPNGYLVLDLSMQ EALSGT	600
	:***.: : **** : **** ** *:***: ***** **:*****:*. : ** *	
Odobenus_rosmarus	AR-LGPG--PVFTAIP TLL LALIPK	399
Mustela_putorius	ARLVGPR--TVLTAIP TLL LALIPK	401
Canis_lupus	ARLLRPG--PVFTAILALLLGLIPK	401
Felis_catus	ARLLRPG--PVLTAIPTLLWALVPN	401
Bos_taurus	PHRLQGPGPVLTVTASLLLVLS	403
Sus_scrofa	SRPLRPGAPVLTLTSLLLALILK	403
Equus_caballus	PRLLGPG--PGLTMILTLLAEILN	401
Homo_sapiens	PCLLGPG--PVLTVLALLLASTLA-	622
	. : . : * ** :	

Mammalia, sequenza completa

Splicing phase		1		0
Exon number		1	2	2 3
Homo sapiens	MALPTARPLLGSCGTPALG-SLLFLLFSL GWVQPSRTLGETGQ EAAPLD			
Pan troglodytes	MALPTARPLLGSCGTPALG-SLLFLLFSF GWVQPSRALAGETGQ EAAPLD			
Nomascus leucogenys	MALLTARSLLGSCGTPALS-SLLLLLFSL GWVQPSRALAGETGQ EAAPLD			
Macaca mulatta	MALPMARPLSGSCGTPALG-SLLFLLFSL GWVQPSRVLAGETGQ EAAPLD			
Papio anubis	MALPTARPLLGSCGTSALG-SLLFLLFSL GWVQPSRALAGETGQ EAAPLD			
Saimiri boliviensis	MALPTARPLWRSCGSPALG-RLPFLLLSL GWVQPWKAWAGETGQ EAAPLD			
Mus musculus	MALPTARPLLGSCGSPICRSFLLLLLSL GWIPRLQTQTTKTSQ EAT-LL			
Rattus norvegicus	MALPTAQPLLGSCGSPICRSFLLLLLSL GWLPLLQTQTTKTSQ EAA-LL			
Cricetulus griseus	MALPTARPLLGSCGSPICSSLLLLLSL GWMPVLQLQATKTDQ EAVHLC			
Odobenus rosmarus	MSLQTARPP LGSGTSAHG-NLLLLLSL GWVFP SRAQAADSRL -----			
Mustela putorius	MAMRTAQPAWGSWGTSAHG-SLLLLLSL GWAFFSRAQAADLRL -----			
Canis lupus	MALQTAQPPLGSRWTSTRG-SLLLLLSL GWVLP SRAQAADSRL -----			
Felis catus	MALQAALPPLGSRGTPHC-SLLLLLSL GWVLP SRVRAADSRP -----			
Bos taurus	MALQTAPPRLGSGWTPAHG-SLLLLLCL GWVLP SRAQAADVGL -----			
Sus scrofa	MALQTARPHLGSRGTRTPR-SLLLLLSL GWVPP SRAQAADTGQ -----			
Equus caballus	MALQTAQRPLGSCGTPTHG-SLLLLLSL GWLPPPGAQAADTGL -----			

Splicing phase	2
Exon number	3 4
Homo sapiens	GVLANPPNISS LSPRQLLGFP CAEVSGLSTERVRELAVALAQKNVKLSTE

Pan troglodytes	GVLANPPNISS LSPRQLLGFPCEVSGLSTERVRELAVALAQKNVKLSAE
Nomascus leucogenys	GVLANAPNIAS LSPRQLLGFRCAEVSGLSTERVQELAVAGQKNVKLSAE
Macaca mulatta	GILTNPADIAS LSPRQLLGFTCEVSGLSTELVQELAVAGQKNVKLSAE
Papio anubis	GVLTNAPDIAS LSPRQLLGFTCEVSGLSTELVQELAVAGQKNVKLSAE
Saimiri boliviensis	RVLANTRNIAS LSPGGLLGFTCAEVSALSPESVWELAVSLGQKNVTLSEAE
Mus musculus	HAVNGAADFAS LPTGLFLGLTCEEVS DLSMEQAKGLAMAVRQKNITLRGH
Rattus norvegicus	HAVTGTVDVAS LPTGLFLGLTCEEVS DLSMGHAKELAMAVRQKNIVLQVH
Cricetulus griseus	DELIGNPDFAS LPAGLFLGLTCEEVS GLSMAHAQALAMAVKQKNITLRVN
Odobenus rosmarus	-----
Mustela putorius	-----
Canis lupus	-----
Felis catus	-----
Bos taurus	-----
Sus scrofa	-----
Equus caballus	-----

Splicing phase	0	2
Exon number	4 5	5 6
Homo sapiens	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTRFFSRITKA	
Pan troglodytes	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTRFFSRVTKA	
Nomascus leucogenys	Q LRCLAHRLSE--PPKDLALPLDLLLFLN PDAFSGPRACTRFFSRVTRA	
Macaca mulatta	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTHFFSRVAKA	
Papio anubis	Q LRCLAHRLSE--PPEDLDALPLDLLLFLN PDAFSGPQACTHFFSRVAKA	
Saimiri boliviensis	Q LRCLAHRLSE--PPEDLDALPPDLLLFLN PDAFSGSQAQACTRFFSRVSKA	
Mus musculus	Q LRCLARRLPKHLTDEELNALPLDLLLFLN PAMFPGQACAHFFSLISKA	
Rattus norvegicus	Q LRCLARRLPKHLTNEELDAPLDLLLFLN PAMFPGQACAHFFSLISKA	
Cricetulus griseus	Q LRCLARRLPKYLNEELDAPLDLLLFFN PAMFRGQACVHFFSLISKA	
Odobenus rosmarus	- - - - -	
Mustela putorius	- - - - -	
Canis lupus	- - - - -	
Felis catus	- - - - -	
Bos taurus	- - - - -	
Sus scrofa	- - - - -	
Equus caballus	- - - - -	

Splicing phase	0
Exon number	6 7
Homo sapiens	NVDLLPRGAPERQRLPAALACW GVRGSLLSEADVRLGGLACDLPGRFV
Pan troglodytes	NVDLLPRGAPERQRLPAALACW GVRGSLLSEADVRLGGLACDLPGRFV
Nomascus leucogenys	NVDLLPREAPERQRLAAALACR GVRGSLLSEADVQALGGLACDLPGRFV
Macaca mulatta	NVDLLPRGAPERQRLPGALTWCW GVRGSLLSEADVRLGGLACDLPGRFV
Papio anubis	NVDLLPRGAPERQRLPAALACW GVQGSLLSEADVQALGGLACDLPGRFV
Saimiri boliviensis	NVDLLPRGAPERRRLLPAALACR GVRGALLSEADVRLGGLACDLPGRFV
Mus musculus	NVDVLPRRSLERQRLMEALKCQ GVGFGVQSEADVRLGGLACDLPGRFV
Rattus norvegicus	NVNVLPRRSLERQRLTGALKCQ GVGFGVQSETDARALGGLACDLPGEFV
Cricetulus griseus	NIDLLPQRSLEQRLPRALECQ GVYSFQLSETDAKALGGLVCDLPGRFV
Odobenus rosmarus	-----
Mustela putorius	-----
Canis lupus	-----
Felis catus	-----

Bos taurus	-----	-----
Sus scrofa	-----	-----
Equus caballus	-----	-----

Splicing phase		2
Exon number		7 8
Homo sapiens	AESAEVLLPRLVSCPGPLDQDQQEAAARAALQGGGPPYG PPSTWSVSTMDA	
Pan troglodytes	AESAEVLLPRLVSCGLGPLDQDQQEAAARAALQGGGPPYG PPSTWSVSTLDA	
Nomascus leucogenys	AESAEVLLPRLARCLGPLDQDQQEAAARAALQGGGPPYG PPSTWSVSTLDA	
Macaca mulatta	AESAEVLLPRLVRCGLGPLDQDQQEAAARAALQGGGPPYG PPSTWSISTLDD	
Papio anubis	AESAEVLLPRLVRCGLGPLDQDQQEAVRAALQGGGPPYG PPSTWSISTLDA	
Saimiri boliviensis	AESAEVLLPRLVHCLGPLDQDQQEAAARAALQGGGPPYG PPSTWSASTLDA	
Mus musculus	ARSSEVLLPWLACGQGPLDQDQQAQAVREVLRSGRTPYQ PPSKWSVSTLDA	
Rattus norvegicus	AKSSEVLLPWLARCGGPLDQDQQAQAVREVLRSGRAPYQ PPSTWSVSTLDA	
Cricetulus griseus	VKSSEVLLPWLACRGLNEGQEEAVREALRSGRSPYQ PPSRWSVSTLNA	
Odobenus rosmarus	-----	
Mustela putorius	-----	
Canis lupus	-----	
Felis catus	-----	
Bos taurus	-----	
Sus scrofa	-----	
Equus caballus	-----	

Splicing phase		0
Exon number		8 9
Homo sapiens	LRGLLPVLGQPIIRSIPQ GIVAARQRSSRDPSWRQPETILRPRFRREV	
Pan troglodytes	LRGLLPVLGQPIIRSIPQ GIVAARQRSSRDPSWQQPERTILWPRFRRDV	
Nomascus leucogenys	LQGLLPVLGRPIIRSIPQ GMVARVAATLPWDPSWRQPEQTXLRPRFRRDV	
Macaca mulatta	LQSLLPVLGQPVISARPQ GILAARQRSSRDPSWQQPEQTVLRLRFRRDV	
Papio anubis	LQSLLPVLGQPVIHSIPQ GILAARQRSSRDPSWRQPEQTIILRPRFRRDV	
Saimiri boliviensis	LRGLLPVLGQPVVRSIPQ GVVAARWKRTSRDPSWRQPETALRPLRRDM	
Mus musculus	LQSLVAVLDESIVQSIPK DVKAEWLQHSRDP SRLGSKLTVIHPRFRDA	
Rattus norvegicus	LQGLLVVLDESIVHSIPK DVITEWLQGISREPSRLGSKWTVTHPRFRDRT	
Cricetulus griseus	LQGLLVVLDESIIHSIPK DVKAEWLQHISRGPSWRGIKLTAMIPRFRDRT	
Odobenus rosmarus	----- GVMTPLWLGTSRDPSWQQPELTVVLLRDRDRT	
Mustela putorius	----- GVMTPLWLGTSRDLVSWRPELSVVLRSRDRDRT	
Canis lupus	----- GVMTPLWLGTSRDLTWQWPELTVILSRDRDRT	
Felis catus	----- GVMTPLWLGTSRDLVSWRPELTVLLPRDRDRT	
Bos taurus	----- ALLSSSVQRTSRDLVSWQPELTVLLPRARRGT	
Sus scrofa	----- GVLIPWLQRTSGDLVSWQPEPTAIRPRPQGGT	
Equus caballus	----- GFVTPWLQRPTRDPSWQQPELTVVLLQRAQRST	

Splicing phase		1
Exon number	9	10
Homo sapiens	E KTACPSGKKAREIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE	
Pan troglodytes	E KTACPSGKKAREIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE	
Nomascus leucogenys	E RTACPPGKEAHEINENLIFYKKWELEACVDAALLAQMDHVNAIPFTYE	
Macaca mulatta	E RTTCPPEKEVHEIDESLIFYKKWELEACVDPALLAQMDRVDAIPFTYE	
Papio anubis	E RTTCPPEKEVHEIDESLIFYKKWELEACVDAALLAQMDRVNAIPFTYE	
Saimiri boliviensis	E KKACPPGKEVLEIDENLIFYKNWELEACVNASLLATQIDRVDAIPFTYE	

Mus musculus	E QKACPPGKEPYKVDEDLIFYQNWLEACVDGTMARQMDLVNEIPFTYE
Rattus norvegicus	E QKACPPGKEPNVVDENLIFYQNWLEACVDGTLLAGQMDLVNEIPFTYE
Cricetulus griseus	E RKACPPGMEPHKVVDENLIFYQNWLEACVDGAMLATQMDLVNEIPFTYE
Odobenus rosmarus	E -KVCPSERKAHVIDENLVFYEEWELEACVDGALLAAQMDRVNMIPFTYQ
Mustela putorius	D SKVCPSPRKAHVIDENLVFYEEWELEACVDGALVAAQMDRVNMVPFTYQ
Canis lupus	Q KKVCPPGQKAHVVDENLVFYEEWELEACVDGALLEAQMDRVNMVPFTYQ
Felis catus	E KKACPPERRARVVDENLVFYEPWELEACVDGALLAAQMDRVNMVPFTYQ
Bos taurus	E KKACPPGREAQVVDENLVFYEEWELEACVDGALLATQMRQVNLIPFTYQ
Sus scrofa	E KRDCPPGRKAQVVDENLVFYEDWELEACVDGALLADQMDQVNLIPFTYE
Equus caballus	K SKACPPEQKATVVDENLVFYEEWELEACVDAALDAQMDRVNAIPFTHQ

Splicing phase	0
Exon number	10 11
Homo sapiens	QLDVLKHKLDE LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLETLKA
Pan troglodytes	QLDVLKHKLDE LYPQGYPESVIQHLGYLFLKMSPEDIRKWNVTSLETLKA
Nomascus leucogenys	QLDVLKHKLDE LYPQGYPESVIQHLGHFLKMSPEDIRKWNVTSLETLKA
Macaca mulatta	QLDVLKHKLDE LYPQGYPESVIRHLGHFLKMSPEDIRKWNVTSLETLKA
Papio anubis	QLDVLKHKLDE LYPQGYPESVIRHLGHFLKMSPEDIRKWNVTSLETLKA
Saimiri boliviensis	QLGILKHKLDE LYPQGYPEPVIQHLGQFFLKMSPEDIRKWNVTSLETLKA
Mus musculus	QLSIFKHKLDK TYPQGYPESLIQQLGHFFRYVSPEDIHQWNVTSPTVKT
Rattus norvegicus	QLSIFKHKLDK TYPQGYPESLIQQLGHFFRYVSPEDIRQWNVTSPTVNT
Cricetulus griseus	QLDIFKHKLDE TYPQGYPESLIQQLGHFFRYISPEDIHRWNVSSPTVKT
Odobenus rosmarus	QLDVFKRKLDQ FYPQGYPESLIQHLRYFFLELTPADIHKWNVTSLETVKS
Mustela putorius	QLDVFKRKLDQ FYPQGYPESLTQNLSYFFLELSPADIRKWNVTSLETVKS
Canis lupus	QLDVFKRKLDQ LYPQGYPESLIQHLGYFFFEITPEDIHKWNVTSLETVKS
Felis catus	QLGVFKRKLDQ LYPQGYPESLVRHLGYFFLELTPEDIHKWNVMSLETVKS
Bos taurus	QLHIFKRLDE FYPQGYPSQLIERLSYFFLYVTPEDIHKWNVTSLETVKS
Sus scrofa	QLHVFKRLDE LYPQGYPTSLVQRLRYFFLSVTPEDICKWNVTSLETVKS
Equus caballus	QLEIFRCKLNQ FYPQGYPESLIRRLRYFFYAFTPEDIYKWNVTSPEIVKY

Splicing phase	0
Exon number	11 12
Homo sapiens	LLEVNGHEMSPQ VATLIDRFVKGQQLDKDITLDTLAFYPGYLCSLSPE
Pan troglodytes	LLEVNGHEMSPQ VATLIDRFVKGQQLDKDITLDTLAFYPGYLCSLSPE
Nomascus leucogenys	LLEVSKGHKMSAQ VATVIDRFVMGRGQLDKDITLDTLAFYPGYLCSLSPE
Macaca mulatta	LLKVSKGHEMSAQ VATLIDRVVGRGQLDKDITVDTLTAFCPGCLCSLSPE
Papio anubis	LLKVSKGHEMSAQ VATLIDRVVGRGQLDKDITVDTLTAFCPGCLCSLSPE
Saimiri boliviensis	LLKVSKGHEMSVQ VATLIDHFLMGRRLDKDITLDTLAFYPGYLCSLSPE
Mus musculus	LLKVSQGQKMQA AIALVACYLRGGGQLDEDMVKALGDIPLSYLCDFSPQ
Rattus norvegicus	LLKVSQGQKMQA VIALVACYLRGGGQLDEDIVKALDNIPLSYLCDFSPQ
Cricetulus griseus	LLKVTKGQKMQA VTALVACYLRGGGQLDQDIVKAMDDIPLTYLCDFKQ
Odobenus rosmarus	LLQVSKGQTMQA VAALIARYVAGGGQLDKASLDTLATFGPTYLCFLSPE
Mustela putorius	LLQVSKGQKRDA VAALIARYVAGGGQLDQATLDMLATFRPAYLCFLSLE
Canis lupus	LLKVSKGHKVDA VAALIARYVAGGGQLDKATLDVLTATFSPTYLCVLSPE
Felis catus	LLEVSKGHEMDAQ VAALIARYLVGGGQLDKATLDTLAAFHPTYLCFLSPE
Bos taurus	LLKVSQGRGVDA VAALIARYVGGGQLDKAALDTLAIFRPAYLCFLRPE
Sus scrofa	LLXVSKGRGMDAQ VAALIARYVGGGQLDGATLDILATFCPTYLCFLSPE
Equus caballus	LLKISEGRMDAQ VAVLIARYLVGGGQMDKATLDTLAAFRPTYLCCLLSPE

Splicing phase	2
----------------	---

Exon number	12	13
Homo sapiens	ELSSVPPSSIW AVRPQDLDTCDPRQLDVLYPKARLAFQNMNGSEYFVKIQ	
Pan troglodytes	ELSSVPPSSIW AVRPQDLDTCDPRQLDVLYPKARLAFQNMNGSEYFVKIQ	
Nomascus leucogenys	QLSSVPPSIIW AIRPQDLDTGPRQLDVLYPKARLAFQNMNGSEYFVKIQ	
Macaca mulatta	RLSSVPPSVIG AVRPQDLDTGPRQLDVLYPKARLAFQNMNGSEYFVKIR	
Papio anubis	RLSMPPSIIG AVRPQDLDTGPRQLDVLYPKARLAFQNMNGSEYFVKIR	
Saimiri boliviensis	QLGSPVPSLIW AVRPQDLDTGPRQLDVLYPKARLAFQNMNGSEYFVKIR	
Mus musculus	DLHSVPSVVMW LVGPQDLDKCSQRHLGLLYQKACSAFQNVSGLEYFEKIK	
Rattus norvegicus	DLHAIPSSVMW LVGLHDLDKCSQRHLGLLYQKACSAFQNVSGLEYFEKIR	
Cricetulus griseus	NLHSIPSSFMW LVKPQDLNCSQRHLGLLYQKACSAFQNVSSRQYEFKIR	
Odobenus rosmarus	QLSSVQPSVLW AARPDLEACRPLQMDVLYPRARAFQNINGSEYFVKIK	
Mustela putorius	QLRSVQHSLVW AAGPQDLDACHPQLQMAVLYPRARTAFQTVSGPEYFSRIK	
Canis lupus	QLGSPVHSLVIW VARAQDLDACRPPQMDVLYPRARTAFQNLGSEYFERIK	
Felis catus	QLESVHHSVVW AARPPDLDACRPAQMDVLYPRARTAFQNLGSEYFEKMK	
Bos taurus	QLDSVQLSVLW MTTPQDLDACSPQMAVLYHKAHMAFQNVSGSEYFTRIK	
Sus scrofa	QLSSIQLSVVR VTRPLDLDTCSPLQVAVLYLKARAFQNMNGSEYLARIK	
Equus caballus	QLDSMPHNVMW AAGPQDLDTCRPPQMDALYSKARIAFQNMNGSEYFARIK	
Splicing phase	1	0
Exon number	13	14
Homo sapiens	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Pan troglodytes	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Nomascus leucogenys	SFL GGAPTEDLKALSQQNASMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Macaca mulatta	PFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Papio anubis	PFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Saimiri boliviensis	SFL GGAPTEDLKALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPHV	
Mus musculus	TFL GGASVKDLRLSQQNVSMDLATFMKLRTDAVL GLSVAEVQKLLGPHI	
Rattus norvegicus	TFL GGASREDLRALSQQNVSMDLATFMKLRTDAVL GLSVAEVQKLLGPHI	
Cricetulus griseus	TFL GGASVEDLRALSQQNVSMDLATFMKLRTDAVL GLTVAEVQKLLGPHI	
Odobenus rosmarus	PYL AGAPTEDLRLSSQNVSMDLATFMKLRTDAVL PLTIAEVQKLLGPNL	
Mustela putorius	PFL GGASTEDLRALSQQNVSMDLATFMKLRTDAVL PLTIAEVQKLLGPNL	
Canis lupus	PYL GGARTEDLRALSQQNVSMDLATFMKLRTDAVL PLTVAEVQKLLGPNL	
Felis catus	LYL GGAPTEDLRLSRQNI SMDLA AFRTLRPEAVV PLTVAEVQKLLGPNL	
Bos taurus	PFL GGASTEDLRVFTSQNI SIDAAAFKLTTEAVL SLTVAEVQKLLGPNL	
Sus scrofa	PYL GGAPTEDLRLSRQNI SLDIATFKKLQTEVLR KLTIAEVQKLLGPHL	
Equus caballus	PYL GGAPTEDLQALSRQNIHMDMATFKKLRTDAVL PLTIAEVQKLLGPNL	
Splicing phase	1	
Exon number	15	16
Homo sapiens	EGLKAEERHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSMQ EAL	
Pan troglodytes	EGLKAEERHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Nomascus leucogenys	EGLKAEERHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Macaca mulatta	EGLKVEEQHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Papio anubis	EGLKVEEQHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Saimiri boliviensis	EGLKAEERHSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Mus musculus	VDLKTEEDKSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Rattus norvegicus	GDLKTEEDKSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Cricetulus griseus	ADLKTEEDKSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVLDLSVR EAL	
Odobenus rosmarus	VGLKAEQESSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVMDLSFR EAL	
Mustela putorius	VGLKAEQERSPVRDWILRQRQDDDLTLGLGLQGGIPNGYLVMDLSFR EAL	

Canis lupus	AGLKAEQGNPVRDWILRQRQDDLRLGLGLRGGIPNGYLVMDLSFR EAL
Felis catus	VGLKAARESSPVRDWISRQRQEDLDSLGLGLRGGIPNGYLVVDLSFR EAL
Bos taurus	VGLKAETGNMFLRDWISRQSQEDLRLGLGLVGGVPNGYLVLDLRGR EAS
Sus scrofa	AGLKAEENSPVRDWISRQPQDDLRLGLGLQGGIPNGYLVLDINIR EAP
Equus caballus	AGLKAEEGNSPVRDWILRQPQDDLNSLGLGLRGGIPNGYLVLDLSFR EAL

Homo sapiens	SGTPCLLGPG--PVLTVLALLLASTLA-
Pan troglodytes	SGTPCLLGTG--PVLTVLALLLASTLA-
Nomascus leucogenys	SGTPCLLGPA--PVLTVLALLLASTLA-
Macaca mulatta	SGTPCLLGPG--PVLTVLALLLASTLA-
Papio anubis	SGTPCLLGPG--PVLTVLALLLASTLA-
Saimiri boliviensis	SGTPCLLGPG--PVLTVLALLLAPTLAC
Mus musculus	SSRASLLGPG--FVLIWIPALLPALRLS
Rattus norvegicus	SSGAPLLGPG--FVFAWIPALLSALRLS
Cricetulus griseus	SSGAPLLGPG--FVIAWIPALLSAFILS
Odobenus rosmarus	SGGAR-LGPG--PVFTAIPITLLALIPK
Mustela putorius	SGGARLVGPR--TVLTAIPTLLALIPK
Canis lupus	SGGARLLRPG--PVFTAILALLGLIPK
Felis catus	SGGARLLRPG--PVLTAIPTLLWALVPN
Bos taurus	SGGPHRLGQGPVLTVTASLLLVLVLS
Sus scrofa	WGGSRPLRPGLAPVLTLTPLSLLALILK
Equus caballus	SGGPRLLGPG--PGLTMILTLLAEILN

G.2 Allineamenti per la sequenza compelta del trascritto

Blocco 1

Homo	ATGGCCTT 176
Pan	ATGGCCTT 170
Nomascus	ATGGCCTT 169
Macaca	ATGGCCTT 8
Papio	ATGGCCTT 166
Saimiri	ATGGCCTT 109
Mus	ATGGCCTT 116
Rattus	ATGGCCTT 98
Cricetulus	ATGGCCTT 8

Homo	GCCAACGGCTCGACCCCTGTTGGGGTCCTGTGGGACCCCGCCCTCG---GCAGCCTCCT 233
Pan	GCCAACGGCTCGACCCCTGTTGGGGTCCTGTGGGACCCCGCCCTCG---GCAGCCTCCT 227
Nomascus	GCTAACGGCTCGATCCCTGTTGGGGTCCTGTGGGACCCCGCCCTCA---GCAGCCTCCT 226
Macaca	GCCAATGGCTCGACCCCTGTCGGGGTCCTGTGGGACCCCGCCCTCG---GCAGCCTCCT 65
Papio	GCCAACGGCTCGACCCCTGTTGGGGTCCTGTGGGACCTCTGCCCTTG---GCAGCCTCCT 223

Saimiri	GCCAACGGCTCGACCCCTGTGGCGGTCTGCGGGAGCCCCGCCCTTG---GCAGGCTGCC	166
Mus	GCCAACAGCTCGACCCCTGTGGGGTCTGTGGAAGTCCCATCTGCAGCCGAAGCTTCCT	176
Rattus	GCCAACAGCCCAACCCCTGTGGGGTCTGTGGAAGCCCATCTGCAGCCGAGCTTTCT	158
Cricetulus	GCCAACAGCTCGACCCCTGTGGGGTCTGTGGGAGCCCATCTGTGCAGTAGCCTTCT	68
	** ** . ** * . * ***** * ***** ** * . * * . * * * * *	
Exon number	1 2	
Homo	GTTCTGTCTTTCAGCCTCG GATGGGTGCAGCCCTCGAGGACCTGGCTGGAGAGACAGG	293
Pan	GTTCTGTCTTTCAGCCTCG GATGGGTGCAGCCCTCGAGGGCCCTGGCTGGAGAGACAGG	287
Nomascus	GTTGCTGTCTTTCAGCCTCG GATGGGTACAGCCCTCGAGGGCCCTGGCTGGAGAGACAGG	286
Macaca	GTTCTGTCTTTCAGCCTCG GATGGGTGCAGCCCTCGAGGGTCTGGCTGGAGAGACAAG	125
Papio	GTTCTGTCTTTCAGCCTCG GATGGGTGCAGCCCTCGAGGGCCCTGGCTGGAGAGACAAG	283
Saimiri	GTTCTGTCTTTCAGCCTCG GATGGGTGCAGCCCTGGAAGGCTTGGGCCGAGAGACAGG	226
Mus	ACTGCTTCTCCTTAGTCTTG GGTGGATACCAGTCTGCAGACCCAGACTACAAAGACAAG	236
Rattus	ACTGCTTCTCCTTAGTCTTG GGTGGTTGCCATTCTGCAGACCCAGACTACAAAGACAAG	218
Cricetulus	ATTACTTCTTCTTAGTCTTG GGTGGATGCCAGTTTTCAGCTCCAGGCTACAAAGACGGA	128
	. * * * * * * * * * * * . * . * * . * . * . * . * . * . * .	
Exon number	2 3 3 4	
Homo	GCAG GAGGCTGCGCCCTGGACGGAGTCTTGCCCAACCCACCTAACATTTCAG CCTCTC	353
Pan	GCAG GAGGCTGCGCCCTGGACGGAGTCTTGCCCAACCCACCTAACATTTCAG CCTCTC	347
Nomascus	GCAG GAGGCTGCGCCCTGGACGGAGTCTTGCCCAACGCACCTAACATTTCAG CCTCTC	346
Macaca	GCAG GAGGCCGCGCCCTGGATGGAATCCTGACCAATGCACCTGACATTGCCAG CCTCTC	182
Papio	GCAG GAGGCCGCGCCCTGGATGGAATCCTGACCAATGCACCTGACATTGCCAG CCTCTC	343
Saimiri	GCAG GAGGCTGCGCCCTGGACCGAGTCTTGCCCAACACGCGTAACATTGCCAG CCTCTC	286
Mus	CCAG GAGGCCACACTCCTCCATGCTG---TGAACGGTGCCGCTGACTTTGCCAG TCTCCC	293
Rattus	CCAG GAGGCCGCACTTCTCCATGCTG---TGACCGGCACCGTTGACTTTGCCAG TCTCCC	275
Cricetulus	CCAG GAGGCTGTACATCTCTGTGATGAACTGATCGGCAACCCCTGACTTCGCCAG TCTCCC	188
	*** ** . * ** . : . ** . * . . * . * : * **** * * *	
Homo	CCCTGCCAACTCCTTGGCTTCCCGTGTGCGGAGGTGTCCGGCCTGAGCACGGAGCGTGT	413
Pan	CCCACGCCAACTCCTTGGCTTCCCATGTGCGGAGGTGTCCGGCCTGAGCACAGAGCGTGT	407
Nomascus	GCCACGCCAACTCCTTGGCTTCCCGTGTGCGGAGGTGTCTGGCCTGAGCACGGAGCGTGT	406
Macaca	CCCACGCCAACTCCTTGGCTTCCCGTGTGCGGAGGTGTCCGGCCTGAGCACAGAGCTCGT	242
Papio	CCCACGCCAACTCCTTGGCTTCCCATGTGTGAGGAGGTGTCTGGCCTGAGCACAGAGCTCGT	403
Saimiri	CCCTGGCGGCCTCCTTGGCTTCCCGTGTGCGGAGGTGTCTGCCTTGAACCCGAGAGTGT	346
Mus	CACAGGCCTCTTCTTGGCCTCACATGTGAGGAGGTATCTGACCTGAGCATGGAACAAGC	353
Rattus	CACAGGCCTCTTCTTGGCCTCACATGTGAGGAGGTATCTGGCCTAAGCATGGGACACGC	335
Cricetulus	TGCAGGCCTCTTCTTGGCCTCACTTGTGAAGAAGTGTCTGGTCTGAGCATGGCACACGC	248
	* : * * . * ***** ** . ***** ** . * . * . * . * . * . * .	
Exon number	4 5	
Homo	CCGGGAGCTGGCTGTGGCCTTGGCACAGAAGAATGTCAAGCTCTCAACAGAGCAG CTGCG	473
Pan	CCGGGAGCTGGCTGTGGCCTTGGCACAGAAGAATGTCAAGCTCTCAGCAGAGCAG CTGCG	467
Nomascus	CCAGGAGCTGGCTGTGGCCTTGGGACAGAAGAACGTCAAGCTCTCAGCAGAGCAG CTGCG	466
Macaca	CCAGGAGCTGGCTGTGGCCTTGGGACAGAAGAATGTCAAGCTCTCCGCAGAGCAG CTGCG	302
Papio	CCAGGAGCTGGCTGTGGCCTTGGGACAGAAGAATGTCAAGCTCTCCGCAGAGCAG CTGCG	463
Saimiri	CTGGGAGCTGGCTGTGCTTGGGACAGAAGAATGTACGCTCTCAGCAGAGCAG CTGCG	406
Mus	CAAGGGGCTGGCTATGGCTGTAAGACAGAAGAATACATCCGGGGACATCAG CTGCG	413
Rattus	CAAGGAGCTGGCTATGGCTGTGAGACAGAAGAATATCGTGCTCCAAGTACATCAG CTGCG	395

Homo	CTGTCTGGCTCACC GGCTCTCTGAGCCCCCG-----AGGACCTGGACGCCCTCCCAT	527
Pan	CTGTCTGGCTCACC GGCTCTCTGAGCCCCCG-----AGGACCTGGATGCCCTCCCAT	521
Nomascus	CTGTCTGGCTCACC GGCTCTCCGAGCCCCCA-----AGGACCTGGACGCCCTCCCGTT	520
Macaca	CTGTCTGGCTCACC AGCTCTCTGAGCCCCCG-----AGGACCTGGACGCCCTCCCGCT	356
Papio	CTGTCTGGCTCACC GGCTCTCTGAGCCCCCG-----AGGACCTGGACGCCCTCCCGCT	517
Saimiri	CTGTCTGGCTCACC GTCTCTCTGAGCCCCCG-----AGGACCTGGATGCCCTCCCACT	460
Mus	TTGTCTGGCAGCTGC GCTTCTTAGGACCTCACCAGCAGGAAGTGAATGCTCTTCCACT	473
Rattus	CTGTCTGGCCCGTGC GCTTCCCTAAGCACCTCACCAACGAGGAACTGATGCTCTCCACT	455
Cricetulus	CTGTCTGGCAGCTGC GCTTCTTAAGTACCTACCAATGAGGAACTGATGCCCTCCCACT	368
	***** * * ** * * * * * ***** ** * * * * *	

Exon number	5	6	
Homo	GGACCTGCTGCTATTCTCTCAA	CCCAGATGCGTTCTCGGGGCCCCAGGCCTGCACCCGTTT	587
Pan	GGACCTGCTGCTATTCTCTCAA	CCCAGATGCGTTCTCGGGGCCCCAGGCCTGCACCCGTTT	581
Nomascus	GGACCTGCTGCTATTCTCTCAA	CCCAGACGCGTTCTCGGGGCCCCGGGCTGCACCCGTTT	580
Macaca	GGACCTGCTGCTCTTCTCTCAA	CCCAGACGCGTTCTCGGGGCCCCAGGCCTGCACCCACTT	416
Papio	GGACCTGCTGCTATTCTCTCAA	CCCAGACACGTTCTCGGGGCCCCAGGCCTGCACCCACTT	577
Saimiri	GGACCTGCTGCTCTTCTCTCAA	CCCGGCCGCGTTCTCAGGGTCCCAGGCCTGCACCCGTTT	520
Mus	GGACCTGCTGCTCTTCTCTCAA	CCCAGCCATGTTTCCAGGGCAACAGGCTTGTGCCCACTT	533
Rattus	GGACCTGCTGCTCTTCTCTCAA	TCCAGCCATGTTTCCGGGGCAACAGGCTTGTGCCCACTT	515
Cricetulus	GGACCTGCTGCTCTTCTTCAA	CCCAGCCATGTTTTCAGGGGCAACAAGCCTGTGTCCACTT	428
	*****	***	*****

Homo	CTTCTCCCGCATCACGAAGGCCAATGTGGACCTGCTCCGAGGGGGGCTCCCGAGCGACA	647
Pan	CTTCTCCCGCGTCACGAAGGCCAATGTGGACCTGCTCCGAGGGGGGCTCCCGAGCGACA	641
Nomascus	CTTCTCCCGCGTAACGAGGGGCCAACGTGGACCTGTTCTCGGGAGGCTCCTGAGCGACA	640
Macaca	CTTCTCCCGCGTCGCGAAGGCCAACGTGGACCTGCTCCGCGGGGGGCTCCTGAGAGACA	476
Papio	CTTCTCCCGCGTCGCGAAGGCCAACGTGGACCTGCTCCGCGGGGGGCTCCTGAGAGACA	637
Saimiri	CTTCTCCCGCGTCTCAAAGGCCAATGTGGACCTGCTCCCTCGGGGGGCTCCTGAGCGACG	580
Mus	CTTCTCCCCTCATCTCTAAAGCCAATGTGGATGTACTCCACGGAGGTCTCTGGAGCGCCA	593
Rattus	CTTCTCCCCTCATCTCTAAAGCCAATGTAAATGTACTCCACGGAGATCTCTGGAGCGCCA	575
Cricetulus	CTTCTCTCTCATCTCTAAAGCCAATATAGATTGCTCCACAGAGGTCTCTGGAGCGCCA	488
	***** * * . * * ***** . , * * ** * * * **** * * * *	

Exon number		6	7	
Homo	GCGGCTGCTGCCTGCGGGCTCTGGCCTGCTGG GGTGTGCGGGGGTCTCTGCTGAGCGAGGC			707
Pan	GCGGCTGCTGCCTGCGGGCTCTGGCCTGCTGG GGTGTGCGGGGGTCTCTGCTGAGCGAGGC			701
Nomascus	GCGGCTGCTGGCTGCGGGCTCTGGCCTGCCGG GGCGTGCGGGGGTCTCTGCTGAGCGAGGC			700
Macaca	GAGGCTGCTGCCCGGGGCTCTGAAC TGCTGG GGTGTGCGGGGGTCTCTGCTGAGCGAGGC			536
Papio	GAGGCTGCTGCCCGGGCTCTGGCCTGCTGG GGTGTGCAAGGGGTCTCTGCTGAGCGAGGC			697
Saimiri	GCGACTGCTGCCCGGGCCCTGGCCTGCCGG GGCGTGCGGGGGGCTCTGCTGAGCGAGGC			640
Mus	GAGGCTGCTGATGGAGGCTCTGAAGTGCCAG GGCGTGATGGATTTC AAGTGAGTGAGGC			653
Rattus	GAGGCTGCTGACCGGGGCTCTGAATGCCAG GGTGTGATGGATTTC AAGTGAGTGAGAC			635
Cricetulus	AAGGCTGCTGCCTAGAGCTCTGGAATGCCAG GGCGTGATAGCTTTCAACTGAGTGAGAC			548
	* * *	*****	* * *	* * * * *
		* * *	* * *	* * * * *

Homo TGATGTGCGGGCTCTGGGAGGCCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCCGAGTC 767

Pan	TGATGTGCGGGCTCTGGGAGGCCTGGCTTGCACCTGCCTGGGCGCTTTGTGGCCGAGTC	761
Nomascus	TGATGTGACAGGCTCTGGGAGGCCTGGCTTGCACCTGCCTGGGCGCTTTGTGGCTGAGTC	760
Macaca	TGATGTACGGGCTCTGGGAGGCCTGGCTTGCACCTGCCTGGGCGCTTTGTGGCCGAGTC	596
Papio	TGATGTACAGGCTCTGGGAGGCCTGGCTTGCACCTGCCTGGGCGCTTTGTGGCCGAGTC	757
Saimiri	CGACGTGCGTGCTCTGGGAGGCCTGGCGTGCACCTGCCCAGGCGCTTTGTGGCCGAGTC	700
Mus	AGATGTGCGGGCTCTCGGAGGCCTGGCTGTGACCTGCCTGGGAAATTTGTGGCCAGATC	713
Rattus	GGATGCACGGGCTCTCGGAGGCCTGGCTGTGACCTGCCTGGGAAATTCGTGGCCAAATC	695
Cricetulus	AGATGCGAAGGCTCTCGGAGGCCTGGCTGTGACCTGCCTGGGAAATATGTGGTCAAATC	608
	** * ... ***** ***** ** ***** ** . *: **** ...**	
Homo	GGCCGAAGTGCTGTACCCCGGCTGGTGAGCTGCCCGGACCCCTGGACCAGGACCAGCA	827
Pan	GGCCGAAGTGCTGTACCCCGGCTGGTGAGCTGCCTGGGACCCCTGGACCAGGACCAGCA	821
Nomascus	AGCAGAAGTGGTGCTACCCCGGCTGGCGCGCTGCCTGGGACCCCTGGACCAGGACCAGCA	820
Macaca	GGCAGAAGTGGTGCTACCCCGGCTGGTCCGCTGCTTGGGACCCCTGGACCAGGACCAGCA	656
Papio	GGCAGAAGTGGTGCTACCCCGGCTGGTGCCTGCCTGGGACCCCTGGACCAGGACCAGCA	817
Saimiri	GGCAGAGGCGGTGCTACCCCGGCTGGTGCACTGCCTGGGACCCCTGGACCAGGACCAGCA	760
Mus	TTCCGAAGTTCTCCTCCCTGGCTGGCAGGATGCCAAGGACCCCTGGACCAGGACCAGGA	773
Rattus	TTCCGAAGTCTCCTCCCTGGCTGGCAAGATGCGGAGGACCCCTGGACCAAGGCCAGGC	755
Cricetulus	TTCCGAAGTCTCCTCCCTGGCTGGCAGGCTGCCGAGGACCCCTGAATGAGGGCCAGGA	668
	* **,* * **,*** ***** ..*** ,*****,* *... *** .	
Exon number	7 8	
Homo	GGAGGCAGCCAGGGCGGCTCTGCAGGGCGGGGACCCCTACGG CCCCCGTCGACATG	887
Pan	GGAGGCAGCCAGGGCGGCTCTGCAGGGCGGGGACCCCTACGG CCCCCGTCGACATG	881
Nomascus	GGAGGCAGCCAGGGCGGCTCTGCAGGGCGGAGGACCCCTACGG CCGCGCTCGACGTG	880
Macaca	GGAAGCAGCCAGGGCGGCTCTGCAGAGAGGAGGACCCCTACGG CCCCCGTCAACGTG	716
Papio	GGAAGCAGTCAGGGCAGCTCTGCAGAGAGGAGGACCCCTACGG CCCCCGTCAACGTG	877
Saimiri	GGAGGCGGCGGGGAGCTCTGCAGGGCGGAGGACCCCTACGG CCCCCGTCGACGTG	820
Mus	AAAGGCAGTCAGGGAGTTCTGAGGAGTGAAGAACCCAATATGG CCCCCATCGAAGTG	833
Rattus	AAAGGCTGTACGGGAGTTCTGAGGAGTGAAGAGCCCTATGG TCCCCATCGACGTG	815
Cricetulus	AGAGGCAGTCAGGGAAGCTCTGAGGAGTGAAGATCCCATACGG CCCCCATCGAGGTG	728
	..*.* * *.***.* *****.* **..** ***..** ** ** **.*.*.* **	
Homo	GTCTGTCTCCACGATGGACGCTCTGCGGGGCTGCTGCCCGTGTGGCCAGCCCATCAT	947
Pan	GTCTGTCTCCACGCTGGACGCTCTGCGGGGCTGCTGCCCGTGTGGCCAGCCCATCAT	941
Nomascus	GTCTGTCTCCACCTGGACGCTCTGCAGGGCCTGCTGCCCGTGTGGCCAGCCCATCAT	940
Macaca	GTCTATCTCCACCTGGACGATCTGCAGAGCCTGTTGCCGTGTGTGGCCAGCCCGTCAT	776
Papio	GTCTATCTCCACCTGGACGCTCTGCAGAGCCTGTTGCCGTGTGTGGCCAGCCCGTCAT	937
Saimiri	GTCAGCCTCCACCTGGACGCTCTGCGGGGCTGCTCCCGTGTGTGGCCAGCCCGTCGT	880
Mus	GTCAGTCTCCACCTGGATGCCCTGCAGAGCTTGGTAGCAGTGTGGATGAGTCCATCGT	893
Rattus	GTCAGTCTCCACCTGGATGCCCTGCAGGGTTTGTGGTAGTGTGGATGAGTCCATTGT	875
Cricetulus	GTCAGTCTCCACCTGAATGCTCTTCAGGGCCTGCTGACAGTGTGGATGAGTCCATCAT	788
	:. ** .**.* * . ** *.*.* ** * *** ** . * **.* *	
Exon number	8 9	
Homo	CCGCAGCATCCCGCAG GGCATCTGG-CCGCGTGGCGGCAACGCTCCTCTCGGGACCCAT	1006
Pan	CCGCAGCATCCCGCAG GGCATCTGG-CCGCGTGGCGGCAACGCTCCTCTCGGGACCCAT	1000
Nomascus	CCGCAGCATCCCGCAG GGCATGGTGGCCCGCGTGGCGGCAACGCTCC-CTTGGGACCCAT	999
Macaca	CTCTGCTCGTCTCTAG GGCATCTGG-CCGCATGGCGGCAACGCTCCTCTCGGGACCCAT	835
Papio	CCACAGCATCCCGCAG GGCATCTGG-CCGCATGGCGGCAACGCTCCTCTCGGGACCCAT	996

Species	Sequence	Position
Saimiri	CCGCAGCATCCCGCAG GGCGTCGTGG-CTGCATGGCGGAAACGCACCTCTCGAGACCCAT	939
Mus	CCAGAGCATCCCCAAG GATGTCAAAG-CTGAATGGCTGCAACACATCTCCAGAGACCCCT	952
Rattus	CCACAGCATCCCTAAG GATGTTATCA-CTGAATGGCTGCAAGGCATCTCCAGAGAGCCCT	934
Cricetulus	CCACAGCATCCCCAAG GATGTCAAAG-CTGAATGGCTGCAACACATCTCCCGTGGCCCT	847
	* . . ** .** * . * : . * * .**** * .** .*: * * * . ** *	
Exon number		9 10
Homo	CCTGGCGGCAGCCTGAACGGACCATCTCTCCGGCCGCGTTCCGGCGGGAAGTGGAGA AGA	1066
Pan	CCTGGCAGCAGCCTGAACGGACCATCTCTTGCCACGGTTCGGCGGGACGTGGAGA AGA	1060
Nomascus	CCTGGCGGCAGCCTGAACAGACTNGTCTTCGGCCGAGGTTCCGGCGGGACGTAGAGA GGA	1059
Macaca	CCTGGCAGCAGCCGGAACAGACCGTCTCTCCGGCTGAGGTTCCGGCGGGACGTGGAGA GGA	895
Papio	CCTGGCGGCAGCCTGAACAGACCATCTCTCCGGCCGAGGTTCCGGCGGGACGTGGAGA GGA	1056
Saimiri	CCTGGCGGCAGCCTGAACGGACCGCCCTCCGTCCGAGGCTCCGGCGGGACATGGAGA AGA	999
Mus	CCAGGCTGGGGTCTAAGCTGACCGTCATACACCCAAGGTTCCGACGGGATGCAGAAC AGA	1012
Rattus	CCAGGCTGGGGTCTAAGTGGACTGTCACACACCCAAGGTTCCGGCGGGACACAGAAC AGA	994
Cricetulus	CCTGGCGGGGATTAAAGCTGACTGCCATGATCCCAAGGTTCCGGCGGGACACAGAGA GGA	907
	** :*** * . * . * . *** . * .** **** .***** . .** . . **	
Homo	CAGCCTGTCTTCAGGCAAGAAGGCCCGCGAGATAGACGAGAGCCTCATCTTCTACAAGA	1126
Pan	CAGCCTGTCCCTCAGGCAAGAAGGCCCGCGAGATAGACGAGAGCCTCATCTTCTACAAGA	1120
Nomascus	CAGCCTGTCCCCCAGGCAAGGAGGCCACGAGATAAACGAGAACCTCATCTTCTACAAGA	1119
Macaca	CAACCTGTCCCCCAGAGAAAGAGGTCCACGAGATAGACGAGAGCCTCATCTTCTACAAGA	955
Papio	CAACCTGTCCCCCAGAGAAAGAGGTCCACGAGATAGACGAGAGCCTCATCTTCTACAAGA	1116
Saimiri	AGGCCTGTCCCCCAGGCAAGGAGGTCTTGAGATAGATGAAAACCTCATCTTCTACAAGA	1059
Mus	AAGCCTGCCCTCCAGGGAAGGAGCCCTACAAGTGGATGAAGACCTCATCTTCTACCAGA	1072
Rattus	AAGCCTGCCCTCCAGGGAAGGAGCCCTAACGTGGTGGATGAAAACCTCATCTTCTACCAGA	1054
Cricetulus	AAGCCTGCCCTCCAGGGAAGGAGCCCAACAAGTGGATGAAAACCTCATCTTCTACCAGA	967
	...**** ** *** .*:..** .:*.*. * **...*****.***	
Homo	AGTGGGAGCTGGAAGCCTGCGTGGATGCGGCCCTGCTGGCCACCCAGATGGACCGCTGA	1186
Pan	AGTGGGAGCTGGAAGCCTGCGTGGATGCGGCCCTGCTGGCCGCCAGATGGACCGCTGA	1180
Nomascus	AGTGGGAGCTGGAAGCCTGCGTGGATGCGGCCCTGCTGGCCGCCAGATGGACCGCTGA	1179
Macaca	AGCGGAGCTGGAAGCCTGCGTGGACCCAGCCCTGCTGGCCGCCAGATGGACCGTGTGG	1015
Papio	AGCGGAGCTGGAAGCCTGCGTGGATGCGGCCCTGCTGGCCGCCAGATGGACCGTGTGA	1176
Saimiri	ATTGGGAGCTGGAAGCTTGTGTGAACGCGTCCCTGCTGGCCACCCAGATAGACCGCTGG	1119
Mus	ATTGGGAGCTGGAGGCTTGTGTAGATGGCACCATGCTGGCCAGACAAATGGACCTTGTGA	1132
Rattus	ATTGGGAGCTGGAGGCTTGTGTGATGGTACCCTGCTGGCCGCCAGATGGACCTTGTGA	1114
Cricetulus	ATTGGGAGCTAGAGCCTGTGTGGATGGTCCATGCTAGCCACCCAGATGGACCTTGTGA	1027
	* ***** .** .** * * * . * ** .**** .*** .** .** .**** * * .	
Exon number		10 11
Homo	ACGCCATCCCTTACCTACGAGCAGCTGGACGTCTCTAAAGCATAAACTGGATGAG CTCT	1246
Pan	ACGCCATCCCTTACCTACGAGCAGCTGGACGTCTCTAAAGCGTAAACTGGATGAG CTCT	1240
Nomascus	ACGCCATCCCTTACCTACGAGCAGCTGGATGTCCTGAAGCATAAACTGGATGAG CTCT	1239
Macaca	ACGCCATCCCTTACCTACGAGCAGCTGGACGTCTCTAAAGCATAAACTGGATGAG CTCT	1075
Papio	ATGCCATCCCTTACCTACGAGCAGCTGGACGTCTCTAAAGCATAAACTGGATGAG CTCT	1236
Saimiri	ACGCCATCCCTTACCTATGAGCAGCTGGGCATCTCTAAAGCACAAGTTGGACGAG CTGT	1179
Mus	ACGAGATTCCTTACCTATGAGCAGCTCAGTATCTTTAAGCATAAACTGGACAAG ACCT	1192
Rattus	ATGAAATTCCTTTACCTACGAGCAGCTCAGCATCTTCAAGCATAAACTGGACAAG ACCT	1174
Cricetulus	ATGAGATTCCTTACCTATGAGCAGCTCGACATCTTCAAGCATAAACTGGACGAG ACCT	108

		* * . ** ***** ***** ***** . . . ** * ***, ** . ****, . ** . *
Homo	ACCCACAAGGTTACCCCGAGTCTGTGATCCAGCACTGGGCTACCTCTTCCTCAAGATGA	1306
Pan	ACCCACAAGGTTACCCCGAGTCTGTGATCCAGCACTGGGCTACCTCTTCCTCAAGATGA	1300
Nomascus	ACCCACAAGGTTACCCCGAGTCTGTGATCCAGCACTGGACCACCTCTTCCTTAAGATGA	1299
Macaca	ACCCACAAGGCTACCCCGAGTCTGTGATCCGGCACTGGGCCACCTCTTCCTCAAGATGA	1135
Papio	ACCCACAAGGCTACCCCGAGTCTGTGATCCGGCACTGGGCCACCTCTTCCTCAAGATGA	1296
Saimiri	ACCCACAGGGCTACCCCGAGCTGTGATCCAGCACTGGGCCAGTTCTTCCTCAAGATGA	1239
Mus	ACCCACAAGGCTATCTGAGTCCCTGATCCAGCAGCTGGGTCACTTCTTCAGATATGTTA	1252
Rattus	ACCCACAAGGCTATCCCGAGTCCCTGATCAAGCAGCTGGGCCACTTCTTCAGATACGTTA	1234
Cricetulus	ACCCACAAGGCTATCTGAGTCCCTGATCCAGCAGCTGGGTCACTTCTTCGATACATCA	1147
	*****, ** ** * ** * ***** . **** ***, * *****, : * . *	

Species	Sequence	Position
Homo	GCCCTGAGGACATTCGCAAGTGAATGTGACGTCCTGGAGACCCCTGAAGGCTTTGCTTG	1366
Pan	GCCCTGAGGACATTCGCAAGTGAATGTGACGTCCTGGAGACCCCTGAAGGCTTTGCTTG	1360
Nomascus	GCCCTGAGGACATTCGCAAGTGAATGTGACGTCCTGGAGACCCCTGAAGGCATTGCTTG	1359
Macaca	GCCCTGAGGACATTCGCAATGAACGTGACGTCCTGGAGACCCCTGAAGGCTCTGCTCA	1195
Papio	GCCCTGAGGACATTCGCAATGAACGTGACGTCCTGGAGACCCCTGAAGGCTCTGCTCA	1356
Saimiri	GCCCTGAGGACATTCGAAAGTGAATGTGACCTCCTGGAGACCTGAAGGCTCTGCTCA	1299
Mus	GCCCTGAAGACATCCACCAAGTGAATGTGACCTCACCAGACACAGTGAAGGCTCTGCTCA	1312
Rattus	GCCCTGAGGACATCCGGCAGTGAATGTGACTTCACCAGACACAGTGAATACTCTGCTTA	1294
Cricetulus	GCCCCGAGGACATCCACCGGTGAATGTGAGCTCTCCAGACACAGTGAAGGCTCTTCTCA	1207

**** *..***** * ..***** ***** ** * ..** ** ..***** :. : * ** .

Exon number	11	12
Homo	AAGTCAACAAAGGGCACGAAATGAGTCCTCAG GTGGCCACCCTGATCAGCCGCTTTGTGA	1426
Pan	AAGTCAACAAAGGGCACGAAATGAGTCCTCAG GTGGCCACCCTGATCAGCCGCTTTGTGA	1420
Nomascus	AAGTCAGCAAAGGGCACAAAAATGAGTGCTCAG GTGGCCACCCTGATTAGCCGCTTTGTGA	1419
Macaca	AAGTCAGCAAGGGGCATGAAATGAGTGCTCAG GTGGCCACCCTGATTAGCCGCTTTGTGG	1255
Papio	AAGTCAGCAAGGGGCATGAAATGAGTGCTCAG GTGGCCACCCTGATTAGCCGCTTTGTGG	1416
Saimiri	AAGTCAGCAAGGGGCATGAAATGAGTGTTTCAG GTGGCCACCCTGATTAGCATTTCCTGA	1359
Mus	AAGTCAGCAAAGGCACAAAAGATGAATGCTCAG GCGATTGCCTTGGTCGCCTGCTATCTTC	1372
Rattus	AAGTCAGCAAAGGCACAAAAGATGGATGCTCAG GTGATTGCCTTGGTCGCCTGCTATCTTC	1354
Cricetulus	AAGTCACCAAAGGCACAAAAGATGGATGCTCAG GTGACTGCCTTGATTGCCTGCTACCTTC	1267
	***** **.***.* *.****.* ***** *.* **.***.* *.* **.* *	

Homo	AGGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACC GCCTTCTACCTGGGT	1486
Pan	AGGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACC GCCTTCTACCTGGGT	1480
Nomascus	TGGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACC GCCTTCTACCCGGGT	1479
Macaca	TGGGAAGGGGCCAGCTAGACAAAGACCGTAGACACGCTGACTGCCTTCTGCCCGGGT	1315
Papio	TGGGAAGGGGCCAGCTAGACAAAGACCGTAGACACGCTGACTGCCTTCTGCCCGGGT	1476
Saimiri	TGGGGAGGAGGCCAGCTGGACAAAGACACCTGGACACCTGACTGCCTTCTACCTGGGT	1419
Mus	GGGGAGGAGGCCAGCTGGACGAGGATATGGTAAAAGCCTGGGCACATCCCGTTAAGCT	1432
Rattus	GGGGAGGAGGCCAAGCTGGACGAGGACATAGTAAAAGCCTGGACAACATCCCCTTAAGTT	1414
Cricetulus	GGGGAGGAGGCCATCTAGACCAGGACATAGTAAAAGCCATGGATGACATCCCTTTAACCT	1327
	*** * * * * ** * * * * * * * * *	

Exon number	12	13
Homo	ACCTGTGCTCCCTCAGCCCCGAGGAGCTGAGCTCCGTGCCCCCAGCAGCATCTG	GGCGG 1546
Pan	ACCTGTGCTCCCTCAGCCCCGAGGAGCTGAGCTCCGTGCCCCCAGCAGCATCTG	GGCGG 1540

Nomascus ACCTATGCTCCCTCAGCCCTGAGCAGCTGAGCTCCGTGCCCCCAGCATCATCTG|GGCGA 1539
 Macaca GCCTGTGCTCCCTCAGCCCCGAGAGGCTGAGCTCCGTGCCCCCAGCGTCATCGG|GGCGG 1375
 Papio GCCTGTGCTCCCTCAGCCCCGAGCGGCTGAGCTCCATGCCCCCAGCATCATCGG|GGCGG 1536
 Saimiri ACCTGTGCTCCCTCAGCCCTGAGCAGCTGGGCTCCGTGCCCCCAGCCTCATCTG|GGCGG 1479
 Mus ATCTATGTGACTTCAGCCCCAGGATCTGCACTCGGTACCTCCAGTGTCTATGTG|GCTGG 1492
 Rattus ACCTATGTGACTTCAGCCCCAGGATCTGCACGCTATACCTCCAGTGTATGTG|GCTGG 1474
 Cricetulus ACTTATGTGACTTCAAACCCAGAACCTGCATCTATACCTCCAGTTTATGTG|GTTGG 1387
 . *.** .* ***. ** . *** .* * *.*** ** ** * *

Homo TCAGGCCCCAGGACCTGGACACGTGTGACCCAAGGCAGCTGGACGTCCTCTATCCCAAGG 1606
 Pan TCAGGCCCCAGGACCTGGACACGTGTGACCCAAGGCAGCTGGACGTCCTCTATCCCAAGG 1600
 Nomascus TCAGGCCCCAGGACCTGGACACGTGTGGCCGAGGCAGCTGGACGTCCTCTATCCCAAGG 1599
 Macaca TCAGGCCCCAGGACCTGGACACGTGTGGCCGAGGCAGCTGGACGTCCTCTATCCCAAGG 1435
 Papio TCAGGCCCCAGGACCTGGACACGTGTGGCCGAGGCAGCTGGACGTCCTCTATCCCAAGG 1596
 Saimiri TCAGGCCCCAGGACCTGGACACGTGTGGCCGAGGCAGCTAGAGGTCCTCTATCCCAAGG 1539
 Mus TTGGGCCCAAGACCTGGACAAGTGACCCAGAGGCATCTGGGTCTCCTCTACCAGAAGG 1552
 Rattus TTGGGCTCCATGACCTGGACAAGTGACCCAGAGGCATCTGGGTATCCTCTATCAGAAGG 1534
 Cricetulus TTAAACCCCAAGACCTGAACAATTGCAGCCAGAGGCATCTGGATATTCTCTATCAAAGG 1447
 * ..* *** ***,** . ** ..* .***** **,* . * ***** *. ****

Homo CCCGCCTTGCTTTCCAGAACATGAACGGGTCCGAATACTTCGTGAAGATCCAGTCCTTCC 1666
 Pan CCCGCCTTGCTTTCCAGAACATGAATGGGTCCGAATACTTCGTGAAGATCCAGTCCTTCC 1660
 Nomascus CCCGCCTTGCTTTCCAGAACATGAATGGGTCCGAATACTTCGTGAAGATCCAGTCCTTCC 1659
 Macaca CCCGCCTTGCTTTCCAGAACATGAGCGGGTCCGAATACTTCGTGAAGATCCGCGCCTTCC 1495
 Papio CCCGCCTTGCTTTCCAGAACATGAGTGGGTCCGAATACTTCGTGAAGATCCGCGCCTTCC 1656
 Saimiri CTCGGCTCGCTTTCCAGAACGTGAATGGGTCCGAATACTTCGTGAAGATCCGGTCCTTCC 1599
 Mus CCTGCTCAGCCTTCCAGAATGTGAGCGGCTAGAATACTTTGAGAAAATCAAGACATTCC 1612
 Rattus CCTGCTCAGCCTTCCAGAACGTGAGCGGCTGGAATACTTTGAGAAAATCAGGACATTTC 1594
 Cricetulus CCTGCTTGCCCTTTCCAGAATTCAAGCAGGCAGGAATACTTTGAGAAAATCAGGACATTCC 1507
 * * ** ** ***** .* . * ***** ***.***. * *.** *

Exon number 13 14
 Homo TGG|GTGGGGCCCCACGGAGGATTTGAAGGCGCTCAGTCAGCAGAATGTGAGCATGGACT 1726
 Pan TGG|GTGGGGCCCCACAGAGGATTTGAAGGCGCTCAGTCAGCAGAATGTGAGCATGGACT 1720
 Nomascus TGG|GTGGGGCCCCACGGAGGATTTGAAGGCTCTCAGTCAGCAGAATGCGAGCATGGACT 1719
 Macaca TGG|GTGGGGCCCCACGGAGGATGTGAAGGCTCTCAGTCAGCAGAATGTGAGCATGGACT 1555
 Papio TGG|GTGGGGCCCCACGGAGGATTTGAAGGCTCTCAGTCAGCAGAATGTGAGCATGGACT 1716
 Saimiri TGG|GTGGAGCCCCGACAGGACTTGAAGCTACTCAGTCAGCGGAATGTGAGCATGGACT 1659
 Mus TGG|GTGGGGCCTCCGTGAAGGACCTGCGGGCCCTCAGCCAGCACAAATGTGAGCATGGACA 1672
 Rattus TGG|GTGGGGCCTCCAGGGAGGACCTGCGGGCCCTCAGCCAGCACAAATGTGAGTATGGACA 1654
 Cricetulus TGG|GTGGGGCCTCTGTGGAGGACCTGAGGGCCCTCAGCCAGCAGAACCTGAGCATGGACT 1567
 *** ***,*** * . ..*** **,. * ***** ** ** ** *****:

Exon number 14 15
 Homo TGGCCACGTTTCATGAAGCTGCGGACGGATGCGGTGCTG|CCGTTGACTGTGGCTGAGGTGC 1786
 Pan TGGCCACGTTTCATGAAGCTGCGGACAGATGCGGTGCTG|CCGTTGACTGTGGCTGAGGTGC 1780
 Nomascus TGGCCACGTTTCATGAAGCTGCGGACAGATGCGGTGCTG|CCGTTGACCGTGGCTGAGGTGC 1779
 Macaca TGGCCACGTTTCATGAAGCTGCGGAGGGAAGCGGTGCTG|CCGTTGACTGTGGCTGAAGTGC 1615
 Papio TGGCCACGTTTCATGAAGCTGCGGAGGGAAGCGGTGCTG|CCGTTGACTGTGGCTGAGGTGC 1776
 Saimiri TGGCCACGTTTCATGAAGCTGCGGACGGAAGCGGTGCTG|CCATTGACCGTGGCTGAGGTGC 1719

Mus	TAGCCACTTTCAAGAGGCTGCAGGTGGATTCCCTGGTG GGGCTGAGTGTGGCTGAGGTAC	1732
Rattus	TAGCCACTTTCAAGAAGCTGCAGGTGGATGCCCTGGTG GGGCTGAGTGTGGCTGAGGTAC	1714
Cricetulus	TAGCCACTTTCAAGAAGCTGCAGGTGGATGCCCTGGTG GGGCTGACTGTGGCTGAGGTAC	1627
	*.***** ***:**.******.*. **: * ** * . *** *****.*.*	
Homo	AGAAACTTCTGGGACCCACGTGGAGGGCCTGAAGGCGGAGGAGCGGCACCGCCCGGTGC	1846
Pan	AGAAACTTCTGGGACCCACGTGGAGGGCCTGAAGGCGGAGGAGCGGCACCGCCCGGTGC	1840
Nomascus	AGAAGCTTCTGGGACCCACGTGGAGGGCCTGAAGGCGGAGGAGCGGCACAGCCCTGTGC	1839
Macaca	AGAAACTTCTGGGACCCACGTGGAGGGCCTGAAGGTGGAGGAGCAGCAGCCCGGTGC	1675
Papio	AGAAACTTCTGGGACCCACGTGGAGGGCCTGAAGGTGGAGGAGCAGCAGCCCGGTGC	1836
Saimiri	AGAAACTTCTGGGACCCACGTGGAGGGCCTGAAGGCGGAGGAGCGGCACAGCCCGGTAC	1779
Mus	AGAAACTTCTGGGGCCAAACATTGTGGACCTGAAGACCGAGGAGGATAAAAGCCCTGTCC	1792
Rattus	AGAAACTTCTAGGGCCACACATTGGGGACCTGAAGACTGAGGAGGATAAAAGCCCTGTCC	1774
Cricetulus	AGAAACTTCTAGGGCCACACATTGCGGATCTGAAGACCGAGGAGACAAAAGCCCTGTCC	1687
	.*.*.*.*.*.* * * * . ***** . ***** . *. . *** * * *	
Homo	GGGACTGGATCCTACGGCAGCGGCAGGACGACCTGGACACGCTGGGGCTGGGGCTACAGG	1906
Pan	GGGACTGGATCCTACGGCAGCGGCAGGACGACCTGGACACGCTGGGGCTGGGGCTACAGG	1900
Nomascus	GGGACTGGATCCTAGGGCAGCGGCAGGACGACCTGGACGCGCTGGGGCTGGGGCTACAGG	1899
Macaca	GGGACTGGATCCTAAAGCAGCGGCAGGACGACCTGGACACACTGGGGCTGGGGCTACAGG	1735
Papio	GGGACTGGATCCTAAAGCAGCGGCAGGACGACCTGGACACACTGGGGCTGGGGCTACAGG	1896
Saimiri	GGGACTGGATCCTACGGCAGCGGCAAGACGACCTGGACACGCTGGGGCTGGGGCTGACAGG	1839
Mus	GTGACTGGCTGTTCCGCAGCATCAGAAAGACCTAGACAGGCTGGGTTTGGGACTTCAGG	1852
Rattus	GGGACTGGCTCTTCCGACAGCAGCAGAAAGACCTGGACAGTCTGGGTTTGGGACTTCAGG	1834
Cricetulus	GGGACTGGCTCTTCCAGCAGCGGCAGGAAGACCTGGACAGGCTGGGTTTGGGACTTCAGG	1747
	* *****.* * . .*****. **..*.******.*.. ***** ****.* ** *	
Exon number	15 16	
Homo	GCGGCATCCCCAACGGCTACCTGGTCCTAGACCTCAGCATGCAAG AGGCCCTCTCGGGGA	1966
Pan	GCGGCATCCCCAACGGCTACCTGGTCCTAGACCTCAGCGTGCGAG AGGCCCTCTCGGGGA	1960
Nomascus	GCGGCATCCCCAACGGCTACCTGGTCCTAGACCTCAGCGTGCGAG AGGCCCTCTCGGGGA	1959
Macaca	GCGGCATCCCCAACGGCTACCTGATCCTAGACCTCAGTGTGCGAG AGGCCCTCTCGGGGA	1795
Papio	GCGGCATCCCCAACGGCTACCTGATCCTAGACCTCAGTGTGCGAG AGGCCCTCTCGGGGA	1956
Saimiri	GCGGCATCCCCAACGGCTACCTCATCCTGGACCGTGGCGTGCGAG AGGCTCTCTCGGGGA	1899
Mus	GTGGCATCCCCAATGGCTACCTGGTCCTGGACTTCAATGTCCGAG AGGCCTTCTCCAGCA	1912
Rattus	GTGGCATCCCCAATGGCTACCTGATCCTAGACTTCAATGTCCGAG AGGCCTTCTCCAGTG	1894
Cricetulus	GTGGCATCCCCAATGGCTACCTGATTCTAGACCTCAATGTCCGAG AGGCCTTCTCCAGTG	1807
	* ***** ***** . * **.*** .. . * *.** ***** **** . * .	
Homo	CGCCCTGCCTCCTAGGACCTGGACCTGTTTCTCACC---TCCTGGCACTGCTCCTAGCCT	2023
Pan	CGCCCTGCCTCCTAGGAACCTGGACCTGTTTCTCACC---TCCTGGCACTGCTCCTAGCCT	2017
Nomascus	CGCCCTGCCTCCTAGGACCTGCACCTGTTTCTCACC---TCCTGGCCCTGCTCCTGGCCT	2016
Macaca	CGCCCTGCCTCCTAGGACCTGGACCTGTACTCACC---TCCTGGCTTTGCTCCTGGCCT	1852
Papio	CGCCCTGCCTCCTAGGACCTGGACCTGTACTCACC---TCCTGGCTTTGCTCCTGGCCT	2013
Saimiri	CGCCCTGCCTGCTAGGGCCCGGACCTGTGCTCACC---TCTTGGCCTTGCTCCTGGCCC	1956
Mus	GAGCCTCACTCCTTGGGCCAGGATTGTGTTTGCATGGATTCCAGCTCTGCTCCCAGCTT	1972
Rattus	GAGCCCCACTCCTTGGGCCAGGATTGTGTTTGCATGGATTCCAGCTCTGCTCTCAGCTT	1954
Cricetulus	GAGCCCCACTTCTTGGACCAGGATTGTGAATTGCATGGATTCCAGCTTTGCTCTCAGCTT	1867
	. ** .** **:*.*.* * * *** * . . * .** ***** .**	

Homo	CCACCCTGGCCTGA
Pan	CCACCCTGGCCTGA
Nomascus	CCACCCTGGCCTGA
Macaca	CCACCCTGGCCTGA
Papio	CCACCCTGGCCTGA
Saimiri	CCACCCTGGCCTGA
Mus	TAAGGCTGAGCTGA
Rattus	TAAGACTGAGCTGA
Cricetulus	TTATACTGAGCTGA
	* *** ****

Blocco 2

Odobenus	ATGTCCTTGCAGACTGCTCGGCCTCCGCTGGGGTCCTGTGGGACCTCTGC	50
Mustela putorius	ATGGCCATGCGAACC GCCACCCGCGCTGGGGTCCTGGGGACCTCCGC	50
Canis	ATGGCCCTGCAGACTGCTCAGCCTCCCTGGGGTCCCGTTGGACCTCCAC	50
Felis catus	ATGGCCTTGCAGGCTGCTCTGCCCTCCCTGGGGTCCCGTGGGACCCCCAC	50
Equus	ATGGCCTTGCAGACTGCTCAACGCCCTGGGGTCCTGCGGGACCCCCAC	50
Bos	ATGGCCTTGCAGACCGCTCCGCCCGCTGGGGTCCTGGGGACCCCCGC	50
Sus	ATGGCCTTGCAGACGGCTCGGCCCCACCTGGGATCCCGGGGACCCGCAC	50
Homo	ATGGCCTTGCCAACGGCTCGACCCCTGTTGGGGTCCTGTGGGACCCCCGC	50
	*** ** *** * * * * * *** ** * * *** *	

Exon number		1	2	
Odobenus	CCACGGCAACCTTCTGCTCCTGCTTCTCAGCCTTG GGTGGGTGTTTCCTT	100		
Mustela putorius	CGGCGGCAGCCTCCTGCTCCTGCTTCTCAGCCTCG GGTGGGCGTTTCCTT	100		
Canis	ACGTGGCAGCCTCCTGCTCCTGCTCCTCAGCCTTG GGTGGGTGTTACCTT	100		
Felis catus	CCACTGTAGCCTCCTGCTCCTGCTTCTCAGCCTTG GGTGGGTGCTACCTT	100		
Equus	CCATGGCAGCCTCCTGCTCTTGCTTCTCAGCCTTG GGTGGCTGCCACCTC	100		
Bos	CCATGGCAGCCTCCTGCTCCTGCTTCTCAGCCTTG GTTGGGTTCACCTT	100		
Sus	TCCTAGAAGCCTCCTGCTCCTGCTTCTCAGCCTCG GGTGGGTGCCACCTT	100		
Homo	CCTCGGCAGCCTCCTGTTCTGCTCTTCAGCCTCG GATGGGTGCAGCCCT	100		
	* * *** ** * **** ** ***** * * *** **			

Exon number		(2	3)	
Odobenus	CCAGGGCCCAGGCTGCAGA-----	-----	119	
Mustela putorius	CCAGGGCCCAGGCTGCAGA-----	-----	119	
Canis	CCAGGGCCCAGGCTGCAGA-----	-----	119	
Felis catus	CCAGGTCGCGGCTGCAGA-----	-----	119	
Equus	CCGGGCCCAGGCTGCAGA-----	-----	119	
Bos	CCAGGGCCCAGGCTGCAGA-----	-----	119	
Sus	CCAGGGCCCAGGCTGCAGA-----	-----	119	
Homo	CGAGGACCTGGCTGGAGAGACAGGGCAG GAGGCTGCGCCCCTGGACGGA	150		
	* ** ** ***** ***			

Exon number		(3	4)
-------------	--	----	----

Odobenus	-----	-----
Mustela putorius	-----	-----
Canis	-----	-----
Felis catus	-----	-----
Equus	-----	-----
Bos	-----	-----
Sus	-----	-----
Homo	GTCTGGCCAACCCACCTAACATTTCAG CCTCTCCCCTCGCCAACTCCT	200

Odobenus	-----	
Mustela putorius	-----	
Canis	-----	
Felis catus	-----	
Equus	-----	
Bos	-----	
Sus	-----	
Homo	TGGCTTCCCGTGTGCGGAGGTGTCCGGCCTGAGCACGGAGCGTGTCCGGG	250

Exon number	(4)	
Odobenus	-----	
Mustela putorius	-----	
Canis	-----	
Felis catus	-----	
Equus	-----	
Bos	-----	
Sus	-----	
Homo	AGCTGGCTGTGGCCTTGGCACAGAAGAATGTCAAGCTCTCAACAGAGCAG	300

Exon number	(5)	
Odobenus	-----	
Mustela putorius	-----	
Canis	-----	
Felis catus	-----	
Equus	-----	
Bos	-----	
Sus	-----	
Homo	CTGCGCTGTCTGGCTCACC GGCTCTCTGAGCCCCCGAGGACCTGGACGC	350

Exon number	(5	6)
Odobenus	-----	-----
Mustela putorius	-----	-----
Canis	-----	-----
Felis catus	-----	-----
Equus	-----	-----
Bos	-----	-----
Sus	-----	-----
Homo	CCTCCATTGGACCTGCTGCTATTCTCTAA CCCAGATGCGTTCTCGGGGC	400

```

Odobenus -----
Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo CCCAGGCCTGCACCCGTTTCTTCTCCCGCATCACGAAGGCCAATGTGGAC 450

```

```

Odobenus -----
Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo CTGCTCCCGAGGGGGGCTCCCGAGCGACAGCGGCTGCTGCCTGCGGCTCT 500

```

```

Exon number          (6   7)
Odobenus -----
Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo GGCTGTGCTGG|GGTGTGCGGGGGTCTCTGCTGAGCGAGGCTGATGTGCGGG 550

```

```

Odobenus -----
Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo CTCTGGGAGGCCTGGCTTGCGACCTGCCTGGGCGCTTTGTGGCCGAGTCG 600

```

```

Exon number          (2)
Odobenus -----CTCGAGACTG|----- 128
Mustela putorius -----CTTGAGACTG|----- 128
Canis -----CTCAAGACTG|----- 128
Felis catus -----CTCAGCACC|----- 128
Equus -----CACAGGACTG|----- 128
Bos -----CGTGGGACTG|----- 128
Sus -----CACGGGACAG|----- 128
Homo GCCGAAGTGCTGTACCCCGGCTGGTGAGCTGCCCGGGACCC CTGGACCA 650

```

* ***

```

Odobenus -----

```



```

Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo -----
Homo GGACCAGCAGGAGGCAGCCAGGGCGGCTCTGCAGGGCGGGGACCCCCCT 700

```

```

Exon number      (7      8)
Odobenus -----
Mustela putorius -----
Canis -----
Felis catus -----
Equus -----
Bos -----
Sus -----
Homo -----
Homo ACGG|CCCCCGTCGACATGGTCTGTCTCCACGATGGACGCTCTGCGGGGC 750

```

```

Exon number      (8      9)
Odobenus -----|GGTGT
Mustela putorius -----|GGTGT 134
Canis -----|GGTGT 134
Felis catus -----|GGTGT 134
Equus -----|GGCTT 134
Bos -----|GCTCT 134
Sus -----|GGCGT 134
Homo -----
Homo CTGCTGCCCGTGTGGGCCAGCCCATCATCCGCAGCATCCCGCAG|GGCAT 800

```

* *

```

Odobenus CATGACCCCATGGCTGCAAGGCACTTCCCGGGACCCATCCTGGCAGCAAC 184
Mustela putorius CATGACCCCATGGCTGCAAGGTACCTCCCGGGTCTTGTCTGGTGGCGAC 184
Canis CATGACCCCATGGCTGCAAGGCACCTCCCGGGACCTGACCTGGCAGTGGC 184
Felis catus CATGACCCCTGGCTGCAAGGCACCTCCCGGGACCTGTCTTGGCAGCGGC 184
Equus CGTGACCCCTGGCTGCAACGCCCCACCCGGGACCCGTCCTGGCAGCAGC 184
Bos TCTGAGCTCAAGTGTGCAGCGCACATCGCGGGACCTGTCTTGGCAGCAGC 184
Sus CCTGATCCCATGGCTGCAGCGCACCTCCGGGGATCTGTCTTGGCAGCCGC 184
Homo CGTGCCCGCTGGCGGCAACGCTCCTCTCGGGACCCATCCTGGCGGCAGC 850

```

* * * * * * * * * * * * * *

```

Exon number      (9      10)
Odobenus CTGAGCTGACTGTCGTCCTCTTGAGGGACCGGCGGGACACGGAGA|---AG 231
Mustela putorius CTGAGCTATCTGTCGTCCTCTCGCGGGACCGGAGGGACACAGACA|GCAAG 234
Canis CTGAGCTCACTGTCATCCTCTCGAGGGACCGGCGGGACACACAGA|AGAAG 234
Felis catus CTGAGCTGACCGTCCTCCTCCCGAGGGACCGGCGGGACACAGAGA|AGAAA 234
Equus CTGAGCTGACTGTCGTCCTCCAGAGGGCCCGAGCGGAGCACAAGA|-GAAG 234
Bos CTGAGCTGACCGTCCTCCTCCCGAGGGCCCGGCGGGACACAGAGA|AGAAG 234
Sus CTGAGCCGACGGCCATCCGTCCAAGGCCTCAGGGGGGCACAGAGA|AGCGG 234
Homo CTGAACGGACCATCCTCCGCGCGCGGTTCCGCGGGAAGTGGAGA|AGACA 900

```

**** * * * * * * * * * *

Odobenus	GTCTGCCCGTCTGAGCGGAAGGCCACGTGATAGATGAAAACCTTGCTTT 281
Mustela putorius	GTCTGCCCGCCCTCGAGGAAGGCCACGTGATAGATGAAAACCTTGCTTT 284
Canis	GTCTGCCCGCCAGGGCAGAAGGCCACGTGGTGGATGAAAACCTTGCTTT 284
Felis catus	GCCTGCCCGCCAGAGCGGAGGGCCCGTGGTGGATGAAAACCTTGCTTT 284
Equus	GCTTGCCCCCGGAGCAGAAGGCCACTGGTGGATGAAAACCTGGTCTT 284
Bos	GCCTGCCCGCCGGGCGGGAGGCCAAGTGGTGGATGAGAACCTCTTCTT 284
Sus	GACTGCCCGCCGGGCGGAAGGCCAGGTGGTGGATGAACACCTCGTGTT 284
Homo	GCCTGTCTTTCAGGCAAGAAGGCCCGGAGATAGACGAGAGCCTCATCTT 950
	* * * * * * * * * * * * * * * *
Odobenus	CTACGAGGAGTGGGAACCTGGAGGCTGTGTGGATGGGGCCTTGCTGGCGG 331
Mustela putorius	CTACGAGGAGTGGGAGCTGGAGGCTGTGTGGATGGGGCCTGGTGGCGG 334
Canis	CTATGAGGAGTGGGAGCTGGAGGCTGTGTGGACGGGGCCTGCTGGAGG 334
Felis catus	CTACGAGCCCTGGGAGCTGGAGGCTGTGTAGATGGGGCCTGCTGGCGG 334
Equus	CTACGAGGAGTGGGAACCTGGAGGCTGTGTGGACGGGGCCTGCTGGACG 334
Bos	CTATGAGGAGTGGGAGCTGGAGGCTGTGTGGACGGGGCCTGCTGGCCA 334
Sus	CTACGAGGATTGGGAGCTGGAGGCTGTGTGGATGGGGCCTGCTGGCGG 334
Homo	CTACAAGAAGTGGGAGCTGGAAGCCTGCGTGGATGCGGCCCTGCTGGCCA 1000
	*** ** ***** ** * * * * * * * * * *
Odobenus	CTCAGATGGACCGCGTGAACATGATACCTTTCACCTACCAGCAGCTGGAT 381
Mustela putorius	CTCAGATGGACCGCGTGAACATGGTGCCCTTCACCTACCAGCAGCTGGAC 384
Canis	CTCAGATGGACCGGGTGAACATGGTGCCCTTCACCTACCAGCAGCTTGAT 384
Felis catus	CTCAGATGGACCGGGTGAACATGGTGCCCTTCACCTACCAGCAGCTGGGC 384
Equus	CTCAGATGGACCGAGTGAACGCGATACCTTCACCCACCAGCAGCTCGAA 384
Bos	CGCAGATGAGACAAGTGAACCTGATTCCTTCACCTACCAGCAGCTGCAC 384
Sus	ATCAGATGGACCGAGTGAACCTGATCCCTTCACCTACGAGCAACTGCAC 384
Homo	CCCAGATGGACCGCGTGAACGCCATCCCTTCACCTACGAGCAGCTGGAC 1050
	***** * ***** * * * * * * * * * *
Exon number	(10 11)
Odobenus	GTTTTTAAGCGCAAACTGGACGAG TTCTACCCACAGGGGTACCTGAGTC 431
Mustela putorius	GTTTTTAAGCGCAAGCTGGACGAG TTCTACCCACAGGGGTACCCGAGTC 434
Canis	GTTTTTAAGCGCAAACTGGACGAG CTCTACCCACAGGGGTACCTGAGTC 434
Felis catus	GTTTTTAAGCGCAAACTGGACGAG CTCTATCCACAGGGGTACCCGAGTC 434
Equus	ATTTTATAGTGCAAAATGAACGAG TTCTACCCACAGGGGTACCCGAGTC 434
Bos	ATTTTCAAGCGCAAACTGGATGAG TTCTACCCGACAGGGGTACCCGAGTC 434
Sus	GTCTTCAAGCTCAAACTGGATGAG CTCTACCCACAGGGGTATCCACCTC 434
Homo	GTCTTAAAGCATAACTGGATGAG CTCTACCCACAAGGTATCCCGAGTC 1100
	* * * * * * * * * * * * * * * *
Odobenus	CCTGATCCAGCATCTGAGGTACTTCTTCCTCGAGCTACCCCCGCGGACA 481
Mustela putorius	CCTGACCCAGAACCTGAGCTACTTCTTCCTCGAGCTACCCCCGCGGACA 484
Canis	CCTGATCCAGCACCTGGGGTACTTTTCTTCGAGATACCCCTGAGGACA 484
Felis catus	CCTGGTCCGGCACCTCGGGTACTTCTTCCTTGAGCTACCCCCGAGGACA 484
Equus	CCTGATCCGGCGCCTACGGTACTTCTTCTATGCGGTACCCCTGAAGACA 484
Bos	CCTGATCGAGCGCCTGAGTTACTTCTTCCTTACGTACCCCTGAGGACA 484
Sus	CCTGGTCCAGCGCCTGAGGTACTTCTTCCTCCTCGGTACCCCGAGGACA 484
Homo	TGTGATCCAGCACCTGGGCTACCTCTTCCTCAAGATGAGCCCTGAGGACA 1150
	* * * * * * * * * * * * * * * *

Odobenus	TCCACAAGTGGAATGTGACATCCCTAGAACTGTGAAATCTCTTCTTCAA	531
Mustela putorius	TCCGCAAGTGGAACGTGACATCGTTGGAAACCGTGAAATCTCTTCTTCAA	534
Canis	TCCACAAGTGGAACGTGACATCCCTGGAAACAGTGAAATCTCTTCTTAAA	534
Felis catus	TCCACAAGTGGAACGTGATGTCCCTGGAAACCGTGAAATCTCTTCTTGAA	534
Equus	TCTACAAGTGGAATGTGACATCCCCGGAGATTGTGAAATATCTGCTTAAA	534
Bos	TCCACAAGTGGAATGTGACTTCCCTGGACACCGTGAAATCTCTGCTCAAA	534
Sus	TCTGCAAGTGGAATGTGACATCCCTGGGAGACTGTGAAATCTCTGCTCANA	534
Homo	TTCCGCAAGTGGAATGTGACGTCCCTGGAGACCCCTGAAGGCTTTGCTTGAA	1200

* * * * * * * * * *

Exon number	(11 12)	
Odobenus	GTCAGCAAAGGACAGACAATGGATGCTCAG GTGGCTGCCCTGATTGCCCG	581
Mustela putorius	GTCAGCAAAGGGCAGAAAGAGGGACGCTCAG GTGGCTGCCCTGATCGCCCG	584
Canis	GTCAGCAAAGGGCAGAAAGTGATGCTCAG GTGGCTGCCCTGATTGCCCTG	584
Felis catus	GTCAGCAAAGGGCAGAAATGGATGCTCAG GTGGCTGCCCTGATTGCCCG	584
Equus	ATCAGCGAGGGGCGCAGAATGGATGCTCAG GTGGCTGCTCTGATTGCCCG	584
Bos	GTCAGCCAAGGGCGTGGTGTGGATGCTCAG GTGGCTGCCCTGATTGCCCG	584
Sus	GTCAGCAAAGGGCGGGGATGGATGCACAG GTGGCCGCCCTGATTGCCCG	584
Homo	GTCAACAAAGGGCAGAAATGAGTCTCAG GTGGCCACCCTGATCGACCG	

*** * * * * * * * * * * * * * * *

Odobenus	CTATGTGGCGGAGGGGGCCAGCTGGACAAGGCCTCTTTGGACACACTGG	631
Mustela putorius	CTATGTGGCGGAGGGGGCAAGCTGGACCAGGCCACCCTGGACATGCTGG	634
Canis	CTACGTGGCAGGAGGGGGCCAACTGGACAAGGCCACCCTGGACGTACTGG	634
Felis catus	CTATCTGGTGGGAGGGGGCCAGCTGGACAAGGCCACCCTGGACACACTGG	634
Equus	CTACCTGGTGGGAGGGGGCCAGATGGACAAGGCCACCCTGGACACACTGG	634
Bos	CTATGTAGGGGAGGAGGCCAGCTGGATAAGGCCGCCCTGGACACCCTGG	634
Sus	CTATGTAGGGGAGGGGGCCAACTAGACGGCGCCACCTTGGACATCCTAG	634
Homo	CTTTGTGAAGGAAGGGGCCAGCTAGACAAAGACACCCTAGACACCCTGA	1300

* * * * * * * * * * * * * * * * *

Odobenus	CCACCTTCGGCCCCACGTACCTGTGTTTCTCAGTCTGAGCAGCTGAGC	681
Mustela putorius	CCACCTTCAGCCCGCATACCTGTGTTTCTCAGTCTGAGCAGCTGCGC	684
Canis	CCACCTTCAGTCCCACGTACCTGTGTCTCAGTCTGAGCAGCTGGGC	684
Felis catus	CCGCTTTCACCTCCGACGTACCTGTGCTTCTCAGTCTGAGCAGCTGGAG	684
Equus	CTGCCTTCGCTCCACCTACTTGTGCCTCCTCAGTCCCGAGCAGCTGGAC	684
Bos	CCATCTTCGGCCCTGCCTACCTGTGCTTCTCAGCCCTGAGCAGCTGGAC	684
Sus	CCACCTTCTGCCCTACCTACCTGTGCTTCTCAGCCCTGAGCAGCTGAGC	684
Homo	CCGCCTTCTACCTGGGTACCTGTGCTCCCTCAGCCCCGAGGAGCTGAGC	1350

* * * * * * * * * * * * * * * * *

Exon number	(12 13)	
Odobenus	TCCGTGCAACCCAGCGTCCTTTG GGCAGCCAGGCCCCAGGACCTGGAGGC	731
Mustela putorius	TCCGTGCAGCACAGCGTCCTTTG GGCAGCCGGGCCCCAGGACCTGGATGC	734
Canis	TCTGTGCAGCACAGTGTATTG GGTGGCCAGGGCCCAGGACCTGGATGC	734
Felis catus	TCAGTGCAACCCAGCGTTGTTG GGCAGCCGGGCCCCGGATCTGGATGC	734
Equus	TCCATGCCACACAACGTGTTG GGCGCCGGGGCCCAGGACCTGGACAC	734
Bos	TCCGTACAGCTCAGCGTCTTTG GATGACCACACCCAGGACCTGGACGC	734
Sus	TCTATACAGCTCAGCGTGTTCG GGTGACCAGGGCCCTGGACCTGGACAC	734

Homo	TCCGTGCCCCCAGCAGCATCTG GGCGGTCAGGCCCCAGGACCTGGACAC 1400
	** * * * * * * * * * * *
Odobenus	ATGCCGCCCACTGCAGATGGACGTCCTCTATCCCCGGGCCCCGCGCGCCT 781
Mustela putorius	ATGCCACCTCTGCAGATGGCCGTCCTCTATCCCCGGGCCCCGCACTGCCT 784
Canis	ATGCCGCCCAACACAGATGGATGTCTCTACCCCAAGGCCCGCACTGCCT 784
Felis catus	ATGCCGTCGGGCCAGATGGACGTACTCTATCGCAAGGCCCGCTTGCCT 784
Equus	GTGCCGCCACCTCAGATGGACGCCCTCTATTCCAAGGCCCGTATCGCCT 784
Bos	GTGCAGCCACCGCAGATGGCTGTCTCTACCACAAGGCCACATGGCCT 784
Sus	GTGCAGCCCCCTGCAGGTGGCTGTCTCTATCTCAAGGCCCGTGTGCCT 784
Homo	GTGTGACCCAAGGCAGCTGGACGTCTCTATCCAAGGCCCGCTTGCTT 1450
	** ** * * * * * * * * * * *
Odobenus	TCCAGAACATCAATGGGTCTGAATACTTCGTGAAGATCAAGCCCTACCTG 831
Mustela putorius	TCCAGACTGTGAGTGGGCTGAGTACTTCTCGAGGATCAAGCCCTTCCTG 834
Canis	TCCAGAACCTGAGCGGGTCTGAATACTTTGAGAGGATCAAGCCCTACCTG 834
Felis catus	TCCAGAACATGAGCGGGTCTGAATACTTCGAGAAGATGAAGCTCTACCTG 834
Equus	TCCAGAACATGAGCGGGTCTGAGTACTTTGCGAGGATCAAGCCCTACCTG 834
Bos	TCCAGAACGTACGCGGGTCTGAATACTTCACAAGGATCAAGCCCTTCCTG 834
Sus	TCCAGAACATGAGTGGGTCTGAATACTTGGCAGGATCAAGCCCTACCTG 834
Homo	TCCAGAACATGAACGGGTCCGAATACTTCGTGAAGATCCAGTCTTCCTG 1500
	***** * * * * * * * * * * * * * * *
Exon number	(13 14)
Odobenus	G CTGGGGCCCCACGGAGGACCTGCGGGCTCTCAGTCGTCGGAAGGTGAA 881
Mustela putorius	G GTGGGGCCTCCACAGAGGACCTGAGGGCTCTCAGTAGTCAGAAGGTGAA 884
Canis	G GTGGGGCCCGCACCGAGGACCTGCGGGCTCTCAGTCAGCAGAACGTAAG 884
Felis catus	G GTGGGGCCCCACGGAGGACCTGCGGGCTCTCAGCCGTCAGAACATAAG 884
Equus	G GTGGAGCCCCACCGAGGACCTGACGGCCCTTAGTCGGCAGAACATACA 884
Bos	G GTGGGGCCTCCACGAAGACCTGCGGGTTTTTACTTCGCAGAACATCAG 884
Sus	G GTGGGGCCCCACGGAGGACCTGCGGGCTCTTAGTCGGCAAAACATAAG 884
Homo	G GTGGGGCCCCACGGAGGATTGAAGGCGCTCAGTCAGCAGAATGTGAG 1550
	* * * * * * * * * * * * * * *
Exon number	(14 15)
Odobenus	CATGGACATGGCCACATTTAAGACGCTGCAGACGGAGGCCGTGCTG CCGC 931
Mustela putorius	CATGGACGTGGCCACATTCCAGAAGCTGCAGACAGAAGCTGTGCTG CCGC 934
Canis	CATGGACACAGCCACGTTCAAGACCCTGCAGAGGGAGGCGGTGCTG CCGC 934
Felis catus	CATGGACTTGGCCGATTCCGGACGCTGCGGCCTGAGGCAGTGGTG CCGC 934
Equus	CATGGACATGGCCACCTTCAAGAAGCTGCGGACAGAAGCTGTGCTG CCGC 934
Bos	TATAGACGCGCGCATTCAGAAGACTGACAACAGAGGCGGTGCTG TCGC 934
Sus	CCTGGACATCGCCACATTCAAGAAGCTGCAGACGGAGGTCTTGC GG AAGC 934
Homo	CATGGACTTGGCCACGTTTCATGAAGCTGCGGACGGATGCGGTGCTG CCGT 1600
	* * * * * * * * * * * * * * *
Odobenus	TGACCATTGCCGAGGTGCAAAATCTTCTGGGTCCAAACCTAGTGGGCCTG 981
Mustela putorius	TGACCATTGCCGAGGTGCAAAACCTTCTGGGTCCAAACCTGGTGGGCCTG 984
Canis	TGACTGTTGCCGAGGTGCAAAACCTTTTGGGTGCAAACTGGCGGGCCTG 984
Felis catus	TGACCGTTGCCGAGGTCCGAAACCTTCTGGGTCCGAACCTGGTGGGCCTG 984
Equus	TGACCATTGCCGAGGTGCAAAACTTCTGGGTCCGAACCTGGCGGGCCTG 984

Bos	TAACCGTGGCCGAGGTGCAGAACTTCTGGGCCAAACCTGGTAGGCCTG	984
Sus	TGACCATCGCTGAGGTGCAGAACTTCTGGGTCCACACCTGGCGGGCCTG	984
Homo	TGACTGTGGCTGAGGTGCAGAACTTCTGGGACCCACGTGGAGGGCCTG	1650
	* * * * *	
Odobenus	AAGGCAGAGCAGGAGAGCAGCCCCGTGCGGGACTGGATCTTACGGCAGCG	1031
Mustela putorius	AAGGCGGAGCAGGAGAGAGCCCCGTGCGGGACTGGATAGTCCGGCAGCA	1034
Canis	AAGGCGGAGCAGGGGAACAGCCCCGTGCGGGACTGGATCTTACGGCAGCG	1034
Felis catus	AAGGCGGCGCGGGAGAGCAGCCCCGTGCGGGATTGGATCTACGGCAGCG	1034
Equus	AAGGCTGAGGAGGGCAACAGCCCCGTGCGGGACTGGATCTCCGGCAGCC	1034
Bos	AAGGCTGAGACAGGAACATGTTCTGCGGGACTGGATATCCCGGCAGTC	1034
Sus	AAGGCGGAGGAGGAACAGCCCTGTGCGGACTGGATCTCCCGGCAGCC	1034
Homo	AAGGCGGAGGAGCGCACCGCCCGGTGCGGGACTGGATCTACGGCAGCG	1700
	***** * * * * *	
Odobenus	GCAGGATGACCTGGACAGTCTGGGCCTGGGGCTCCACGGTGGCATCCCCA	1081
Mustela putorius	GCAGGATGACCTGGACAGTCTGGGCTTGGGGCTGCACGGTGGCTCCCCA	1084
Canis	GCAGGATGACCTGGACCGTCTGGGCCTGGGGCTCCGCGGTGGCATCCCCA	1084
Felis catus	GCAGGAGGACCTGGACAGTCTGGGCCTGGGGCTCCGAGCGGCATCCCCA	1084
Equus	GCAGGACGACCTGAACAGTCTGGGTCTGGGTCTCCGCGGCGCATCCCCA	1084
Bos	GCAGGAGGACCTGGACAGGCTGGGGCTGGGGCTCGTCGGCGGCGTCCCCA	1084
Sus	GCAGGACGACCTGGACCGGCTGGGGCTGGGGCTCCAGGGCGGCATCCCCA	1084
Homo	GCAGGACGACCTGGACACGCTGGGGCTGGGGCTACAGGGCGGCATCCCCA	1750
	***** ***** * * * * *	
Exon number	(15 16)	
Odobenus	ATGGCTACCTGGTCATGGACCTTAGCTTCCGAG AGGCCCTCTCAGGGGGC	
Mustela putorius	ATGGCTACCTGGTCATGGACCTCAGGCTCCGAG AGGCCCTCTCAGGGGGC	1134
Canis	ACGGCTATCTGGTCATGGACCTCAGCTTCCGAG AGGCCCTCTCAGGGGGC	1134
Felis catus	ATGGCTACCTGGTCGTGGACCTCAGCTTCCGAG AGGCCCTCTCAGGGGGC	1134
Equus	ACGGCTACCTGGTCCTGGACCTCAGCTTCCGAG AGGCCCTCTCAGGGGGC	1134
Bos	ACGGCTACCTGGTCCTGGACCTCCGCGGACGAG AGGCCCTCTCAGGGAGGC	1134
Sus	ACGGCTACCTGGTCCTGGACATCAACATCCGAG AGGCCCTCTCAGGGAGGC	1134
Homo	ACGGCTACCTGGTCCTAGACCTCAGCATGCAAG AGGCCCTCTCAGGGGACG	1800
	* * * * * * * * * * * * * * * *	
Odobenus	GCCCCCTTG---GAC-----CCGACCAGTGTTCACTGCAATCCCCAC	1172
Mustela putorius	GCCCCCTTGTGGGAC-----CCAGAACAGTGCTCACTGCGATCCCCAC	1178
Canis	GCCCCCTCCTCCGAC-----CGGACCAGTGTTCAACGCCATCCTGGC	1178
Felis catus	GCCCCCTGCTCCGAC-----CCGACCTGTGCTCACTGCGGTCCCGAC	1178
Equus	CCCCGCCTCCTTGGAC-----CTGGACCTGGGCTCACTATGATCCTGAC	1178
Bos	CCCCACCGCTCGGACAAGGGCCGGGACCCGTGCTCACTGTGACCGCAAG	1184
Sus	TCCCCGCCCTCAGACCAGGACTTGACCTGTGCTCACGCTGACCCCAAG	1184
Homo	CCCTGCCTCCTAGGAC-----CTGGACCTGTCTCACCG---TCCTGGC	1841
	* * * * * * * * * * * * * * *	
Odobenus	TCTGCTCCTGGCTTTGATCCCCAAGTGA	1200
Mustela putorius	TCTGCTCCTGGCTCTGATCCCCAAGTGA	1206
Canis	TCTGCTCCTGGGTTTGAATCCCAAGTGA	1206
Felis catus	TCTGCTGTGGGCTTTGGTCCCAAGTGA	1206

Equus	TCTGCTCCTGGCTGAGATCCTGAACTGA	1206
Bos	TCTGCTCCTGGICTTGGTCCTGAGCTGA	1212
Sus	TCTGCTCCTGGCCTTGATCCTGAAATGA	1212
Homo	ACTGCTCCTAGCCTCCACCCTGGCCTGA	1869
	***** * ** ***	

Il riferimento numerico degli esoni si riferisce alla sequenza di *Homo sapiens*.

Calcolo dei valori del Likelihood Ratio Test per i risultati dell'analisi con PAML

Per determinare la significatività dei valori di omega ricavati dalla comparazione dei risultati dei diversi modelli analizzati con PAML, si esegue un test di rapporto di verosimiglianza (LRT likelihood ratio tests) che si calcola come il doppio del logaritmo nella differenza di valori verosimiglianza ($\ln L$) tra i due modelli che vengono comparati, ossia $2\Delta\ln L$. Nello specifico, sono stati comparati un modello nullo dove ω non può essere >1 ed un modello alternativo in cui può.

Per il calcolo del LRT, bisogna prendere il valore logaritmico di verosimiglianza per ciascun modello (restituito nel file di output), dove $\ln L1$ corrisponderà al modello alternativo ed $\ln L0$ al modello nullo, e quindi si calcola LRT secondo la formula: $2(\ln L1 - \ln L0)$. Il doppio del $\Delta\ln$ segue una curva di distribuzione χ^2 con 1 grado di libertà, per la quale è possibile ricavare il p value, anche se viene suggerito di comparare il test statistico con una commistione 50:50 del point mass 0 e del χ^2_1 .

I test condotti sono tra:

1. M2a-M1a: il modello M1a di quasi neutralità viene confrontato con il modello M2a di selezione positiva;
2. M8-M7: il modello 7 è definito come modello β a due parametri liberi (p e q), e si può assimilare ad un modello di quasi neutralità, mentre il modello 8 ($\beta&\omega$) ha quattro parametri liberi, p_0 , p , q ed $\omega_s > 1$, e può essere considerato come un modello di selezione positiva. Entrambi seguono una distribuzione di tipo β .
3. M8-M8a: il modello M8a è analogo al modello M8, ma pone $\omega_s = 1$.

4. M3-M0: il modello M0 è quello si base dell'analisi condotta da PAML, in cui l'unico parametro libero è ω ; il modello M3 invece è definito "discreto", ed assume 5 parametri liberi, tra cui ω_0 , ω_1 ed ω_2 . Questo test è suggerito per valutare la variabilità di ω tra siti piuttostochè come test di selezione positiva.

Il valore di p-value viene calcolato come distribuzione χ del valore di LRT, in accordo ai gradi di libertà per ciascun test (1 per il test M2a-M1a, M8-M7, M8-M8a, e 4 per il test M3-M0).

H.0.1 LRT M2a-M1a

MSLN con outgroup				
	M2a	M1a	LRT	p-val
E13	-1,055	-1,058	7,42129	0,00644
E15	-1,407	-1,407	0,00038	0,98441

(a) LRT M2a-M1a con outgroup

MSLN senza outgroup				
	M2a	M1a	LRT	p-val
E13	-911	-913	5,51287	0,01888
E15	-1,144	-1,144	0	1

(b) LRT M2a-M1a senza outgroup

MSLN senza Chimpanzee				
	M2a	M1a	LRT	p-val
E12	-1,279	-1,279	0	1
E13	-1,052	-1,056	7,63861	0,00571
E15	-1,391	-1,391	0,01372	0,90675

(c) LRT M2a-M1a senza Pan troglodytes

Tabella H.1: LRT M2a-M1a

H.0.2 LRT M8-M7

MSLN con outgroup				
	M8	M7	LRT	p-val
E13	-1,054	-1,058	6,65535	0,00988
E15	-1,407	-1,415	16,59219	4,63414E-05

(a) LRT M8-M7 con outgroup

MSLN senza outgroup				
	M8	M7	LRT	p-val
E13	-909	-913	6,29959	0,01208
E15	-1,145	-1,146	2,27999	0,13105

(b) LRT M8-M7 senza outgroup

MSLN senza Chimpanzee				
	M8	M7	LRT	p-val
E12	-1,277	-1,278	1,80844	0,17869
E13	-1,051	-1,055	6,85608	0,00883
E15	-1,391	-1,399	16,17727	5,76822E-05

(c) LRT M8-M7 senza Pan troglodytes

Tabella H.2: LRT M8-M7

H.0.3 LRT M8-M8a

MSLN con outgroup				
	M8	M8a	LRT	p-val
E13	-1,054	-1,057	6,34661	0,01176
E15	-1,407	-1,407	0	1

(a) LRT M8-M8a con outgroup

MSLN senza outgroup				
	M8	M8a	LRT	p-val
E13	-909	-912	5,68997	0,01706
E15	-1,145	-1,145	0	1

(b) LRT M8-M8a senza outgroup

MSLN senza outgroup				
	M8	M8a	LRT	p-val
E12	-1,277	-1,277	0,74971	0,38657
E13	-1,051	-1,055	6,60236	0,01018
E15	-1,391	-1,391	0,02877	0,86531

(c) LRT M8-M8a senza Pan troglodytes

Tabella H.3: LRT M8-M8a

In tutti questi test per la presenza di selezione positiva, solo l'esone 13 è risultato positivo, mentre gli esoni 12 e 15, per quanto presi singolarmente al loro interno pre-

sentino dei siti soggetti a selezione positiva, nel complesso non risultano essere andati incontro, nel corso della loro evoluzione ad accelerazione.

H.0.4 LRT M3-M0

MSLN con outgroup				
	M0	M3	LRT	p-val
E13	-1,086	-1,053	65,56244	1,95872E-13
E15	-1,438	-1,407	62,48147	8,72419E-13

(a) LRT M3-M0 con outgroup

MSLN senza outgroup				
	M0	M3	LRT	p-val
E13	-932	-909	44,66380	2,33943E-11
E15	-1,179	-1,144	68,81468	1,08165E-16

(b) LRT M3-M0 senza outgroup

MSLN senza outgroup				
	M0	M3	LRT	p-val
E12	-1,292	-1,276	32,37782	1,26927E-08
E13	-1,084	-1,051	65,88118	4,78944E-16
E15	-1,421	-1,390	61,60457	4,19843E-15

(c) LRT M3-M0 senza Pan troglodytes

Tabella H.4: LRT M3-M0

Dall'analisi, sembrerebbe che il valore di omega tra siti cambi significativamente negli esoni presi in analisi, ma bisogna ricordare che nel loro complesso, a parte l'esone 13, questi non risultano sottoposti a selezione positiva, mantenendosi quindi evolutivamente stabili.

Alberi filogenetici generati con Mega5

I.1 Analisi della sequenza aminoacidica; E1–16

Blocco 1; sequenza proteica intera, senza outgroup (fig. I.1).

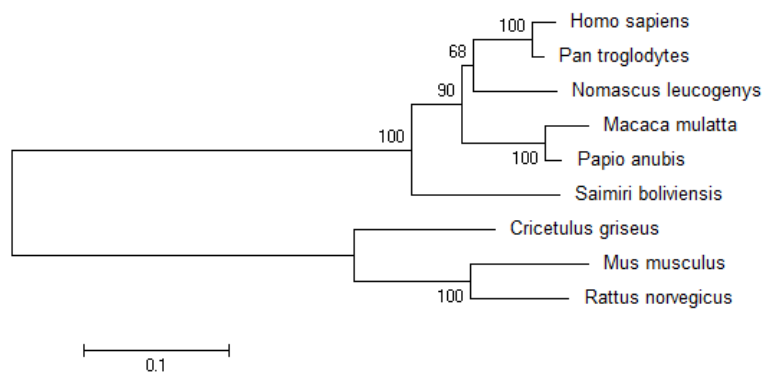


Figura I.1: Albero di maximum likelihood per il blocco 1, analizzando la sequenza proteica intera, senza l'outgroup

Essendo un albero senza radice che ha alla base una dicotomia con la posizione del criceto non risolta, la radice teoricamente può essere collocata in corrispondenza di uno qualsiasi dei rami, modificando la topologia, e quindi l'albero che è possibile ottenere. La ricostruzione ottenuta rispetta nondimeno la filogenesi di riferimento tracciata sulla base dell'intero genoma, con il cluster dei roditori che si separa alla base da quello dei primati (100%), come per i nodi interni, che riconoscono correttamente un cluster di primati composto da *Nomascus leucogenys*, *Pan troglodytes* e *Homo sapiens* (Apes), che fa da sister group ad uno con le Cercopithecinae *Macaca mulatta* (Papionini) e *Pa-*

papio anubis (Baboons), che nel complesso appartengono ai Catarrhini, mentre *Saimiri boliviensis* (Squirrel monkey; New world monkey) si colloca su una diramazione indipendente, essendo una “scimmia del nuovo mondo” (Platyrrhini).

Blocco 1; sequenza proteica intera, con outgroup (fig. I.2).

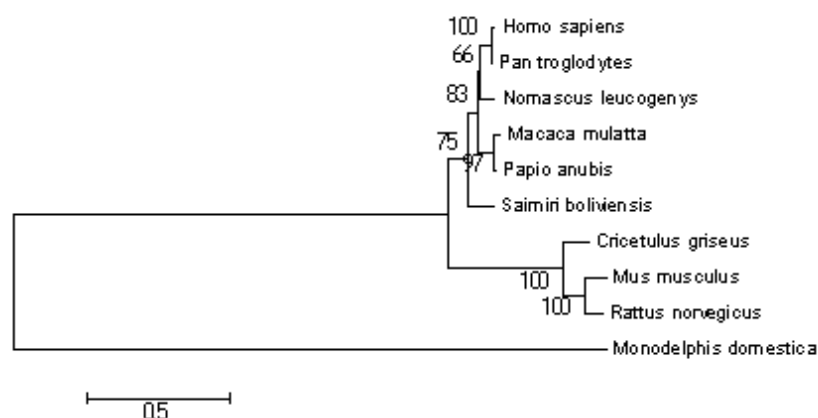


Figura I.2: Albero di maximum likelihood per il blocco 1, analizzando la sequenza proteica intera, con l’outgroup

L’aggiunta dell’outgroup porta ad una riduzione della solidità generale dei nodi, che rimangono nondimeno sopra al 50% di significatività, in particolare per le prime divergenze nel gruppo dei primati, mentre la distinzione alla base tra roditori e primati, e le terminazioni dell’albero, rimangono confermati con valore di bootstrap più affidabili. La filogenesi viene comunque confermata.

Blocco 2; sequenza proteica intera, senza outgroup (fig. I.3).

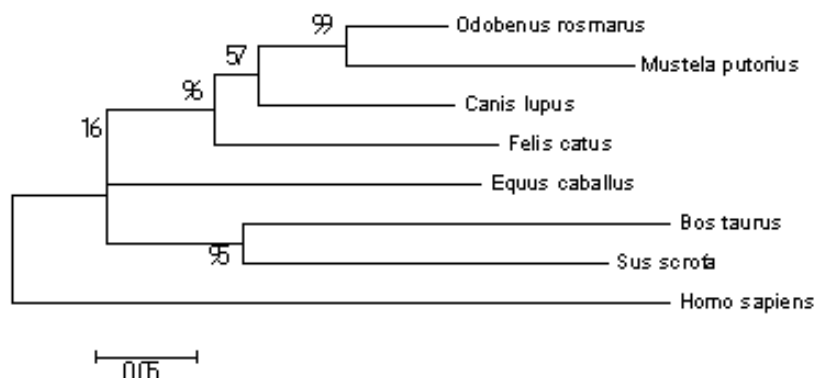


Figura I.3: Albero di maximum likelihood per il blocco 2, analizzando la sequenza proteica intera, senza l'outgroup

Passando ai gruppo dei Laurasiateri, si può osservare come a livello proteico la definizione dell'albero filogenetico sia più povera; solo i nodi che separano *Felis Catus* dai Caniformia (*Canis lupus familiaris*, *Mustela putorius furo* e *Odobenus rosmarus divergens*) e questi ultimi al loro interno, mostrano una significatività apprezzabile, nonché quello che pone *Bos taurus* e *Sus scrofa* come sister groups. L'uomo viene correttamente discriminato e posto come outgroup dei Laurasiateri, ma non risulta rimane la posizione di *Equus caballus*, ed infatti si osserva una tricotomia tra questo, Carnivora ed Artiodactyla.

Se si guarda alla storia evolutiva del gruppo Laurasiateria, si possono distinguere due eventi principali di radiazione: quello che ha portato ai Carnivora (superordine Ferae), e quello che ha portato invece agli ordini di Insectivora, Chiroptera, Cetartiodactyla e Perissodactyla. I Carnivora includono correttamente *Canis lupus* e *Felis catus*, ma anche i mustelidi (*Mustela putorius furo*), che rappresentano di fatto una sottodivisione dei Caniformia, proprio come *Odobenus rosmarus divergens* (Pinnipedia). *Equus caballus* rappresenta, tra le specie da me scelte, il solo rappresentante dei perissodattili, e dovrebbe rappresentare il sister group degli artiodattili *Sus scrofa* e *Bos taurus*. Questo da un punto di vista molecolare-paleontologico, ma si è già sottolineato come la storia evolutiva di un singolo gene non debba ricalcare necessariamente la storia evolutiva delle specie, motivo per cui almeno per la sequenza aminoacidica si può dire che la mesotelina ha una genealogia differente dalla filogenesi delle specie.

La mancata corrispondenza potrebbe dipendere anche da dati insufficienti tra artiodattili e perissodattili per definire correttamente i rapporti evolutivi.

Blocco 2; sequenza proteica intera, con outgroup (fig. I.4).

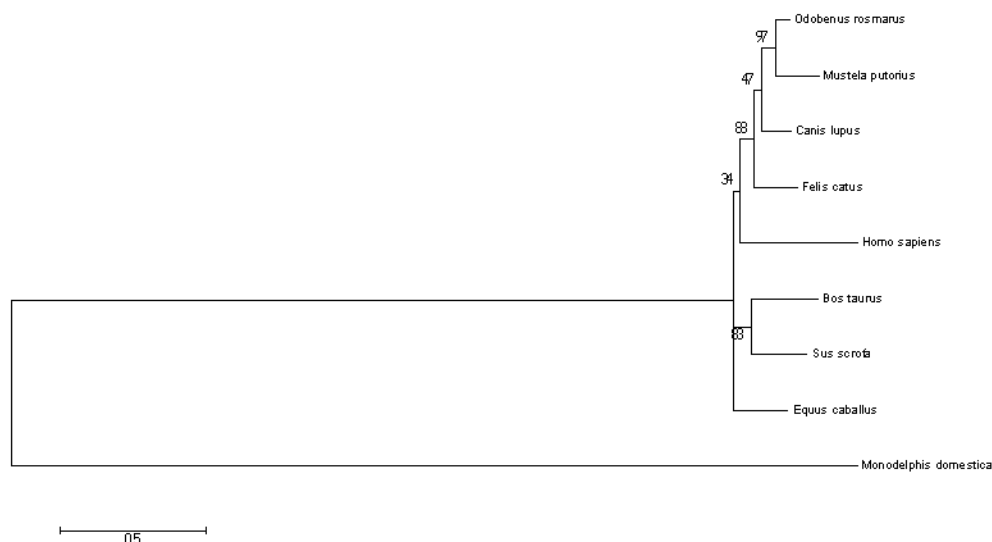


Figura I.4: Albero di maximum likelihood per il blocco 2, analizzando la sequenza proteica intera, con l'outgroup

Quando si inserisce l'outgroup, l'albero assume una topologia differente, ma ancora sono presenti nodi che non raggiungono il supporto statistico necessario a poter trarre conclusioni; viene qui ricostruita una filogenesi in accordo con quella delle specie. É possibile distinguere il cluster dei Carnivora da quello degli Artiodattili-Perissodattili, ma manca il relativo supporto del bootstrap proprio perchè in corrispondenza del nodo potrebbe anche collocarsi la radice, ed *Equus caballus* rappresenta comunque un nodo a parte in quella che è una separazione tricotomica. *Homo sapiens* viene qui posto all'interno del gruppo dei Laurasiateri, come sister group del clade dei Carnivora, per quanto con una affidabilità del 34% che dà poco credito alla ricostruzione.

I.2 Analisi della sequenza del cDNA; E1-16

Blocco 1; cDNA per intero, senza outgroup (fig. I.5).

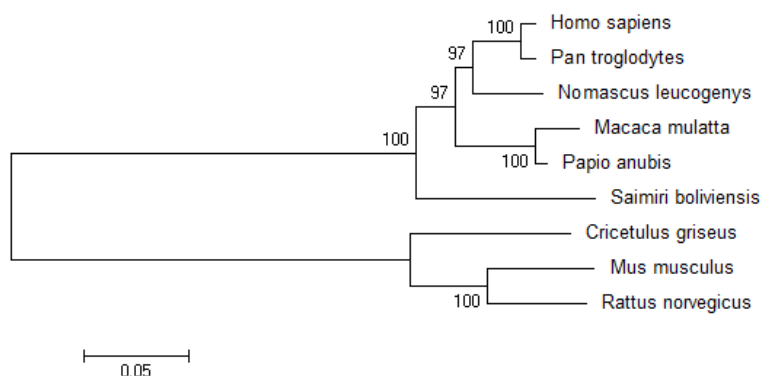


Figura I.5: Albero di maximum likelihood per il blocco 1, analizzando la sequenza del cDNA per intero, senza l'outgroup

Ricordando che l'albero è senza radice, per quanto riguarda il cluster Homo-Pan-Nomascus (Apes), la filogenesi è mantenuta e sostenuta da un valore di bootstrap pari al 97%; questo è poi distinto rispetto al cluster delle Cercopithecinae (Papio-Macaca), con il quale vanno a definire il gruppo dei Catarrhini sister dei Platyrrhini.

Blocco 1; cDNA per intero, con outgroup (fig. I.6 nella pagina successiva).

L'aggiunta dell'outgroup non porta a grandi cambiamenti nell'albero, ed infatti i soli nodi che non raggiungono una solidità del 100% sono quelli relativi alla separazione del *Nomascus* e delle Cercopithecinae. *Cricetulus griseus* viene qui ora saldamente inserito nel cluster dei roditori con una affidabilità del 100%, ed anche tutti gli altri nodi sono nel complesso robusti, ricalcando la filogenesi delle specie. L'albero può quindi essere considerato come un valido strumento di partenza per analisi successive.

Blocco 2; cDNA per intero, senza outgroup (fig. I.7 nella pagina seguente). La filogenesi viene rispettata, dal momento che *Homo sapiens* emerge come primo taxa alla base dell'albero, comportandosi da outgroup, e si possono riconoscere i due grandi cladi di Carnivora ed Artiodattili. Solo *Equus caballus* discosta dalla ricostruzione filogenetica, andando ad inserirsi assieme ai Carnivora e non come sister group degli Artiodattili; questa volta si ha un supporto dell'82%, e viste le precedenti ricostruzioni si può pen-

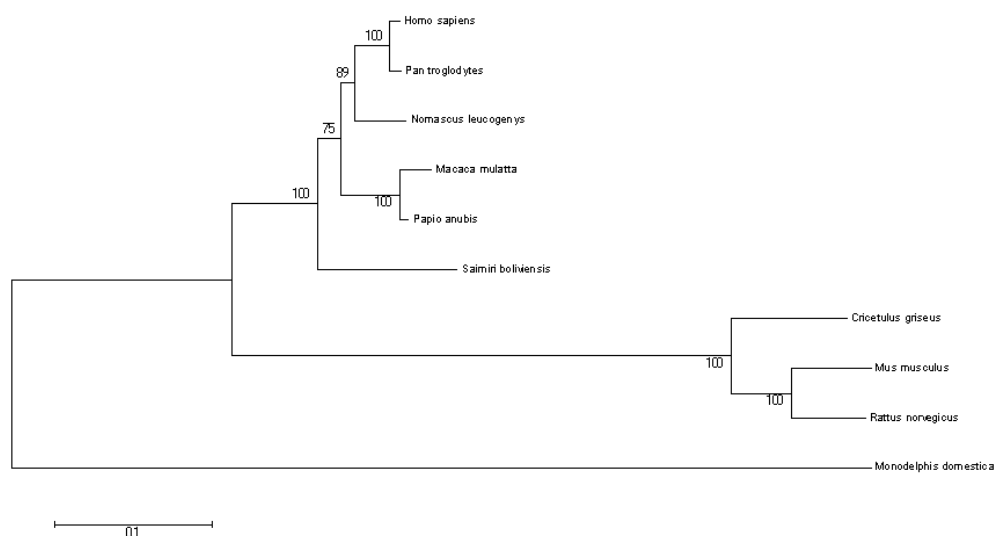


Figura I.6: Albero di maximum likelihood per il blocco 1, analizzando la sequenza del cDNA per intero, con l'outgroup

sare che effettivamente per il gene *MSLN* la storia evolutiva è differente rispetto alla filogenesi.

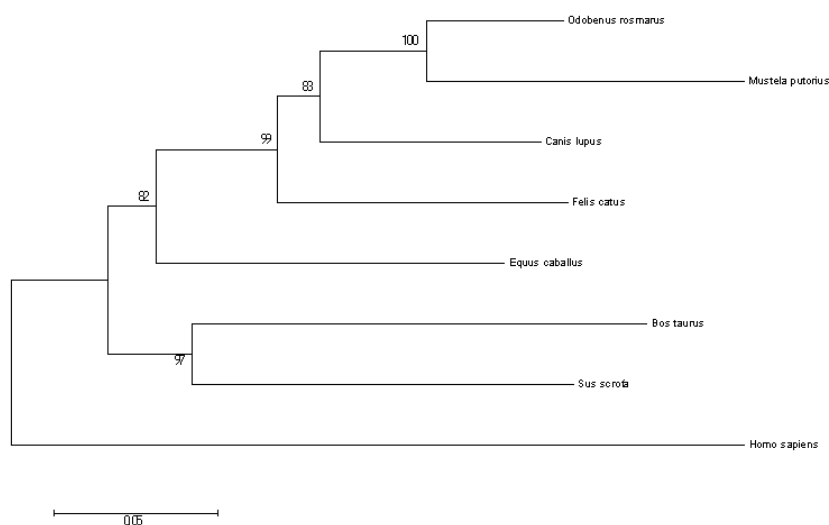


Figura I.7: Albero di maximum likelihood per il blocco 2, analizzando la sequenza del cDNA per intero, senza l'outgroup

Blocco 2; cDNA per intero, con outgroup (fig. I.8).

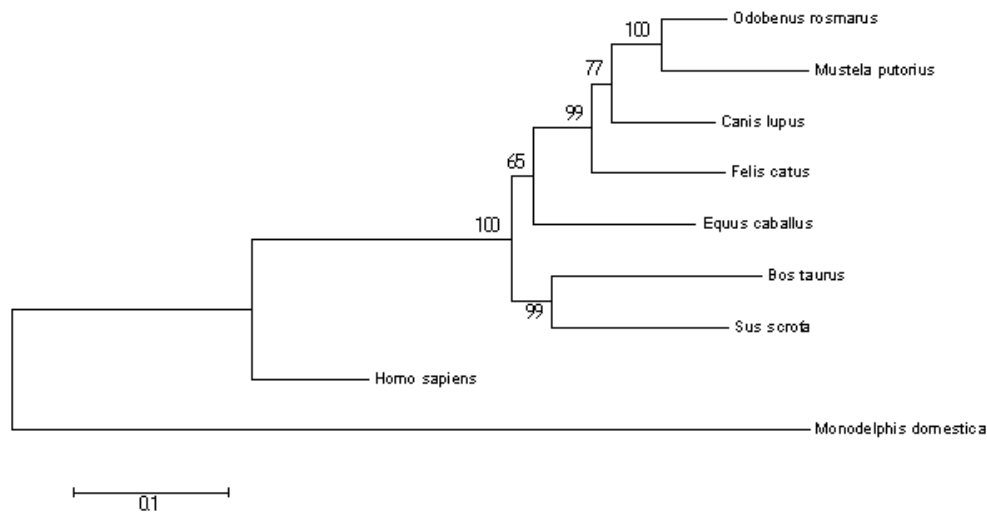


Figura I.8: Albero di maximum likelihood per il blocco 2, analizzando la sequenza del cDNA per intero, con l'outgroup

L'aggiunta dell'outgroup porta a pochi cambiamenti nella topologia dell'albero, con la filogenesi che viene confermata e supportata da alti valori di bootstrap. Rispetto all'analogo ottenuto sulla base della sequenza proteica, *Homo sapiens* è posto all'esterno del cluster dei Laurasiateri, in corrispondenza del cui nodo potrebbe anche collocarsi la radice.

Analisi statistica completa per valutare l'effetto del trattamento con microRNA sul messaggero della mesotelina

Dal momento che anche il tempo sperimentale di estrazione si è dimostrato un fattore sperimentale significativamente positivo, è stata eseguita una analisi statistica tempo per tempo ed esperimento per esperimento, in cui i valori “puri” di mesotelina ricavati dalla normalizzazione per l'espressione genica mediante real-time RT-PCR sono stati analizzati in relazione al trattamento ed all'esperimento, ed al trattamento ed al tempo, rispettivamente. Si è escluso dall'analisi l'esperimento 3, poichè per un errore nel conteggio cellulare sono state seminate più cellule, ed è quindi cambiato il rapporto di espressione del gene.

J.1 Analisi statistica tempo per tempo

J.1.1 ANOVA 2 fattori, 24h

Interaction Plot Nell'interaction plot è possibile osservare come a seconda dei diversi esperimenti sia poco variata l'espressione della mesotelina tra casi (linea rossa) e controlli (linea blu) nelle estrazioni fatte a 24h dalla trasfezione. Sono stati utilizzati i valori “puri” della mesotelina normalizzata per gli housekeeping.

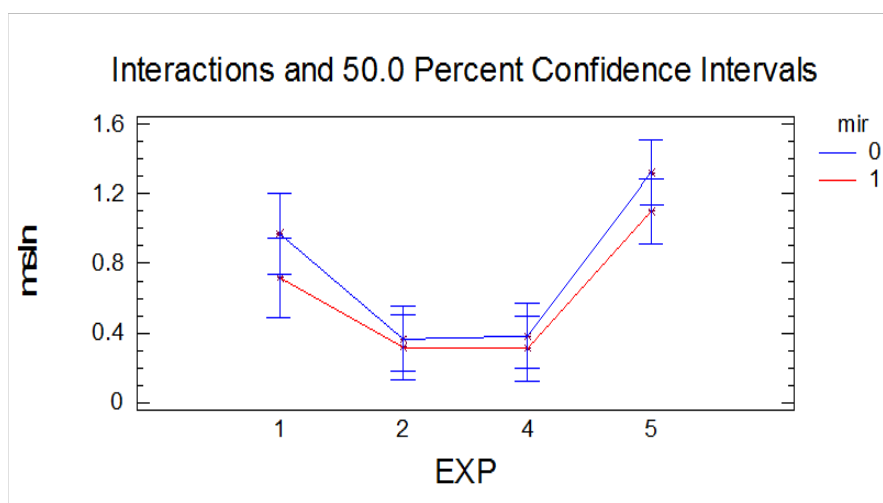


Figura J.1: Interaction plot tra i valori di espressione della mesotelina alle 24h in funzione all'esperimento ed al trattamento. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: Exp	3,18578	3	1,06193	4,79	0,0169
B: mir	0,11798	1	0,11798	0,53	0,4778
Interactions					
AB	0,04341	3	0,01447	0,07	0,9774
RESIDUAL	3,10431	14	0,22174		
TOTAL (corrected)	6,44027	21			

Tabella J.1: Analisi della varianza per i valori della mesotelina a 24h dalla trasfezione

Se nella analisi con i valori di mesotelina normalizzati come Z-score si era eliminata la variabile “esperimento” dal contesto, quando l'esperimento viene trattato come una variabile sperimentale si ottiene una significatività statistica della sua influenza nel far variare i valori di mesotelina, ma il fatto che l'interazione Esperimento-Trattamento non è significativa indica che il disegno sperimentale è comunque valido, poichè i singoli esperimenti indipendenti sono riproducibili.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	22	0,68595			
EXP					
1	4	0,84472	0,23544	0,68170	1,00775
2	6	0,33926	0,19224	0,20615	0,47237
4	6	0,34571	0,19224	0,21260	0,47882
5	6	1,21412	0,19224	1,08101	1,34723
mir					
0	11	0,76032	0,14418	0,66049	0,86015
1	11	0,61159	0,14418	0,51175	0,71142
EXP by mir					
1; 0	2	0,97084	0,33297	0,74028	1,20139
1; 1	2	0,71861	0,33297	0,48806	0,94917
2; 0	3	0,36183	0,27187	0,17358	0,55007
2; 1	3	0,31670	0,27187	0,12845	0,50495
4; 0	3	0,38237	0,27187	0,19412	0,57061
4; 1	3	0,30905	0,27187	0,12080	0,49729
5; 0	3	1,32625	0,27187	1,13801	1,51450
5; 1	3	1,10199	0,27187	0,91374	1,29024

Tabella J.2: Tabella dei least squares means per la mesotelina a 24h dalla trasfezione, con intervalli di confidenza del 50%

J.1.2 ANOVA 2 fattori, 48h

Interaction Plot

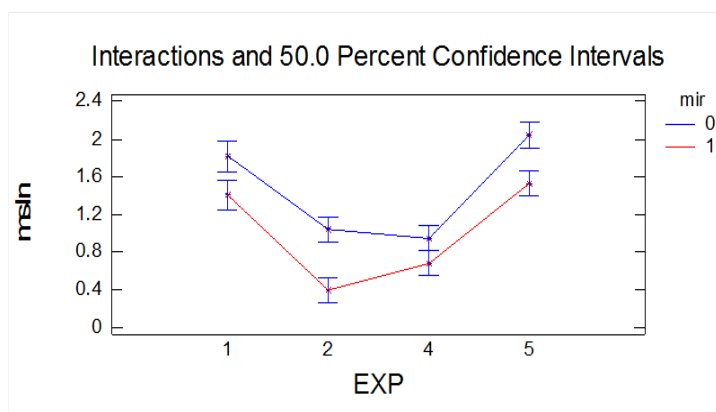


Figura J.2: Interaction plot tra i valori di espressione della mesotelina alle 48h in funzione all'esperimento ed al trattamento. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: Exp	5,03805	3	1,67935	15,16	0,0001
B: mir	1,11634	1	1,11634	10,07	0,0068
Interactions					
AB	0,10939	3	0,03646	0,33	0,8044
RESIDUAL	1,55131	14	0,11081		
TOTAL (corrected)	7,8701	21			

Tabella J.3: Analisi della varianza per i valori della mesotelina a 48h dalla trasfezione

Alle 48h dalla trasfezione si apprezza l'effetto maggiore del trattamento con il microRNA, poichè nei campioni trattati si ha un significativo abbassamento nei livelli di espressione del messaggero della mesotelina (p-value= 0,0068), come osservato anche nell'analisi con i valori di espressione normalizzati per lo Z-score. Che l'effetto si veda 48h dopo la trasfezione è compatibile con i tempi di internalizzazione del microRNA, la trascrizione del gene e l'esportazione del messaggero al citoplasma, dove si ha il ruolo regolatorio negativo del microRNA a livello anche di de-stabilizzazione e degradazione del messaggero. Anche il fattore "esperimento" risulta significativo, ma che l'interazione tra i due fattori sperimentali in analisi non lo sia indica che gli effetti osservati non sono dovuti al caso, ma la regolazione post-trascrizionale è indipendente dal singolo esperimento.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	22	1,23037			
EXP					
1	4	1,61107	0,16644	1,49582	1,72631
2	6	0,71648	0,13590	0,62239	0,81058
4	6	0,80987	0,13590	0,71578	0,90397
5	6	1,78404	0,13590	1,68995	1,87814
mir					
0	11	1,45912	0,10192	1,38855	1,52969
1	11	1,00161	0,10192	0,93104	1,07219
EXP by mir					
1; 0	2	1,81792	0,23538	1,65494	1,9809
1; 1	2	1,40421	0,23538	1,24123	1,56719
2; 0	3	1,03699	0,19219	0,90392	1,17007
2; 1	3	0,39598	0,19219	0,26290	0,52905
4; 0	3	0,94425	0,19219	0,81118	1,07733
4; 1	3	0,67550	0,19219	0,54242	0,80857
5; 0	3	2,03732	0,19219	1,90425	2,17040
5; 1	3	1,10199	0,271868	0,91374	1,29024

Tabella J.4: Tabella dei least squares means per la mesotelina a 48h dalla trasfezione, con intervalli di confidenza del 50%

J.1.3 ANOVA 2 fattori, 72h

Solo gli esperimenti 1, 2 e 3 sono stati analizzati fino alle 72h dalla trasfezione, e si riportano quindi i risultati per tutte e tre le repliche sperimentali indipendenti condotte.

Interaction Plot

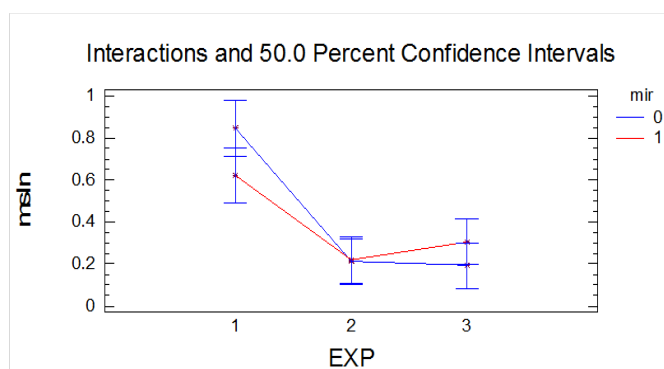


Figura J.3: Interaction plot tra i valori di espressione della mesotelina alle 72h in funzione all'esperimento ed al trattamento. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: Exp	0,75470	2	0,37735	5,21	0,0282
B: mir	0,00474	1	0,00474	0,07	0,8033
Interactions					
AB	0,06889	2	0,03444	0,48	0,6351
RESIDUAL	0,72484	10	0,07248		
TOTAL (corrected)	1,54895	15			

Tabella J.5: Analisi della varianza per i valori della mesotelina a 72h dalla trasfezione

A 72h dalla trasfezione non è più possibile rilevare alcun effetto del trattamento con il microRNA, e solo l'esperimento rimane una variabile statisticamente significativa, per quanto comunque si mantenga indipendente dal trattamento.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	16	0,39927			
EXP					
1	4	0,73287	0,13461	0,63867	0,82708
2	6	0,21542	0,10991	0,13850	0,29234
3	6	0,24952	0,10991	0,17260	0,32644
mir					
0	8	0,41680	0,09693	0,34897	0,48464
1	8	0,38174	0,09693	0,31391	0,44958
EXP by mir					
1; 0	2	0,84523	0,19037	0,71201	0,97846
1; 1	2	0,62051	0,19037	0,48729	0,75374
2; 0	3	0,21166	0,15544	0,10288	0,32044
2; 1	3	0,21918	0,15544	0,11040	0,32796
3; 0	3	0,19351	0,15544	0,08473	0,30229
3; 1	3	0,30553	0,15544	0,19675	0,41431

Tabella J.6: Tabella dei least squares means per la mesotelina a 72h dalla trasfezione, con intervalli di confidenza del 50%

J.2 Analisi statistica esperimento per esperimento

Vengono infine riportati i risultati dell'analisi statistica per i livelli di espressione del messaggero della mesotelina normalizzato per gli housekeeping valutando singolarmente gli esperimenti indipendenti, così da ricavare l'andamento, o meglio la cinetica di espressione, nel tempo.

J.2.1 ANOVA 2 fattori, Esperimento 1

Interaction Plot

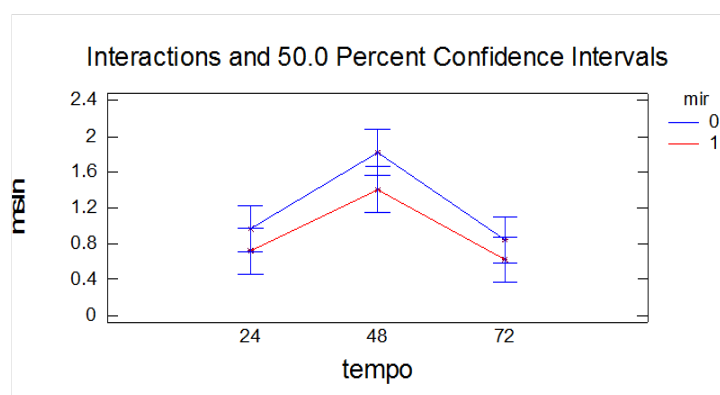


Figura J.4: Interaction plot tra i valori di espressione della mesotelina in funzione tempo e del trattamento per l'esperimento 1. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: mir	0,26442	1	0,26442	1,04	0,3475
B: tempo	1,82801	2	0,91401	3,59	0,0943
Interactions					
AB	0,02085	2	0,01042	0,04	0,9601
RESIDUAL	1,52757	6	0,25459		
TOTAL (corrected)	3,64086	11			

Tabella J.7: Analisi della varianza per i valori della mesotelina registrati per l'esperimento 1

Nel primo esperimento l'effetto del trattamento non è risultato significativo, ed in generale il set di cellule trasfettato mostra livelli di espressione un poco inferiore al set

di cellule di controllo. Nonostante questo, si può apprezzare la variazione cinetica di espressione del gene, che è bassa alle 24h, più alta alle 48h dalla trasfezione, e nuovamente bassa alle 72h, a supportare quanto sottolineato nella discussione al paragrafo 4.4 a pagina 65 riguardo l'apparente cinetica di espressione del gene.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	12	1,06289			
mir					
0	6	1,21133	0,20599	1,06352	1,35914
1	6	0,91445	0,20599	0,76663	1,06226
tempo					
24	4	0,84472	0,25229	0,66369	1,02576
48	4	1,61107	0,25229	1,43003	1,79210
72	4	0,73287	0,25229	0,55184	0,913906
mir by tempo					
0; 24	2	0,97084	0,35679	0,71482	1,22685
0; 48	2	1,81792	0,35679	1,56190	2,07394
0; 72	2	0,84523	0,35679	0,58922	1,10125
1; 24	2	0,71861	0,35679	0,46259	0,97463
1; 48	2	1,40421	0,35679	1,14819	1,66023
1; 72	2	0,62051	0,56788	0,36450	0,87653

Tabella J.8: Tabella dei least squares means per la mesotelina nell'esperimento 1, con intervalli di confidenza del 50%

L'esperimento 1, a differenza di quelli successivi, era stato condotto con repliche sperimentali in doppietta, e con un tempo di estrazione in più alle 3h dalla trasfezione; dal momento però che non si ricavava alcun dato informativo, si è scelto di procedere con le estrazioni unicamente a 24, 48 e 72 ore dalla trasfezione, e facendo tre repliche sperimentali per condizione del trattamento.

J.2.2 ANOVA 2 fattori, Esperimento 2

Interaction Plot

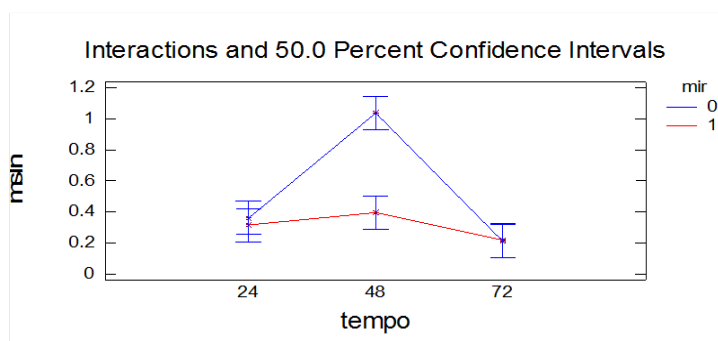


Figura J.5: Interaction plot tra i valori di espressione della mesotelina in funzione tempo e del trattamento per l'esperimento 2. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: tempo	0,81740	2	0,40870	5,60	0,0191
B: mir	0,23027	1	0,23027	3,16	0.1009
Interactions					
AB	0,38922	2	0,19461	2,67	0.1100
RESIDUAL	0,87519	12	0,07293		
TOTAL (corrected)	2,31208	17			

Tabella J.9: Analisi della varianza per i valori della mesotelina registrati per l'esperimento 2

Nel secondo esperimento l'effetto del trattamento non è risultato significativo, ma il fattore tempo lo è, ed infatti quello che si registra alle 48h è un decremento molto significativo dei livelli di espressione nelle cellule trasfettate con il microRNA sintetico. La cinetica sembra qui ulteriormente confermata, e l'assenza di una variazione sostanziale tra casi e controlli alle 24 ed alle 72 ore dalla trasfezione è comprensibile se si considera che a questi intervalli i livelli di messaggero sono talmente bassi che la presenza o l'assenza del microRNA non è rilevante all'alterazione della sua espressione. Quando invece la mesotelina è molto espressa, allora si può apprezzare l'effetto del trattamento.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	12	1,06289			
tempo					
24	6	0,33926	0,11025	0,26258	0,41594
48	6	0,71648	0,11025	0,63981	0,79316
72	6	0,21542	0,11025	0,13874	0,29210
mir					
0	9	0,53683	0,09002	0,47422	0,59943
1	9	0,31062	0,09002	0,24801	0,37323
tempo by mir					
24; 0	3	0,36183	0,15592	0,25339	0,47027
24; 1	3	0,31670	0,15592	0,20826	0,42514
48; 0	3	1,03699	0,15592	0,92855	1,14543
48; 1	3	0,39598	0,15592	0,28754	0,50442
72; 0	3	0,21166	0,15592	0,10322	0,32010
72; 1	3	0,21918	0,15592	0,11074	0,32762

Tabella J.10: Tabella dei least squares means per la mesotelina nell'esperimento 2, con intervalli di confidenza del 50%

J.2.3 ANOVA 2 fattori, Esperimento 3

Interaction Plot

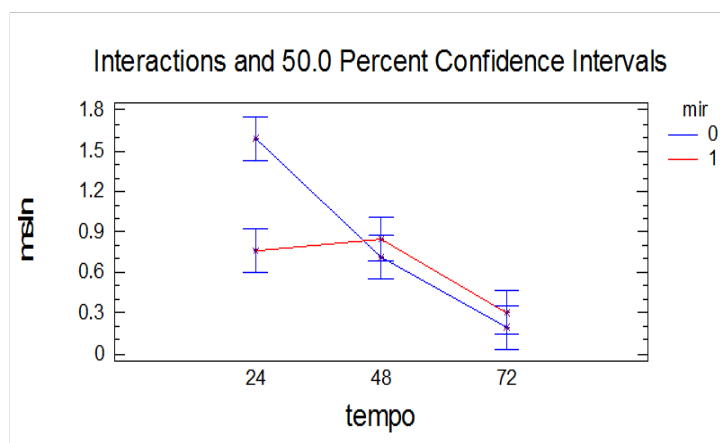


Figura J.6: Interaction plot tra i valori di espressione della mesotelina in funzione tempo e del trattamento per l'esperimento 2. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Anche nel terzo esperimento l'effetto del trattamento non è risultato significativo, mentre il fattore tempo lo è; in questo esperimento, per un errore in fase di conteggio delle

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: tempo	2,59596	2	1,29798	8,17	0,0058
B: mir	0,16883	1	0,16883	1,06	0,3229
Interactions					
AB	0,89863	2	0,44932	2,83	0,0985
RESIDUAL	1,90632	12	0,15886		
TOTAL (corrected)	5,56974	17			

Tabella J.11: Analisi della varianza per i valori della mesotelina registrati per l'esperimento 3

cellule o di semina, sono state seminate più cellule, e questo lo si è visto da subito al momento della trasfezione, poichè nelle piastre vi erano molte più cellule che nelle prove precedenti. Al momento delle estrazioni, è stato come se i livelli di confluenza delle 24h fossero come quelli delle 48h, e così via per le estrazioni successive. Se a livello di analisi statistica questo ha fatto sì che l'esperimento 3 andasse escluso dai conteggi, a livello interpretativo sembrerebbe invece osservare il trend di cinetica già osservato, con un picco di espressione del gene quando le cellule sono al massimo dell'attività proliferativa, ed una sua riduzione quando si raggiunge la piena confluenza della piastra (già alle 48 ore dall trasfezione, ed ancor più alle 72h).

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	18	0,73534			
tempo					
24	6	1,17653	0,16272	1,06336	1,28970
48	6	0,77997	0,16272	0,66680	0,89313
72	6	0,24952	0,16272	0,13635	0,362687
mir					
0	9	0,83219	0,13286	0,73979	0,92459
1	9	0,63849	0,13286	0,54609	0,73089
tempo by mir					
24; 0	3	1,58931	0,23012	1,42927	1,74935
24; 1	3	0,76375	0,23012	0,60371	0,92379
48; 0	3	0,71374	0,23012	0,55370	0,87378
48; 1	3	0,84619	0,23012	0,68615	1,00623
72; 0	3	0,19351	0,23012	0,03347	0,35355
72; 1	3	0,30553	0,23012	0,14549	0,46557

Tabella J.12: Tabella dei least squares means per la mesotelina nell'esperimento 3, con intervalli di confidenza del 50%

J.2.4 ANOVA 2 fattori, Esperimento 4

Interaction Plot

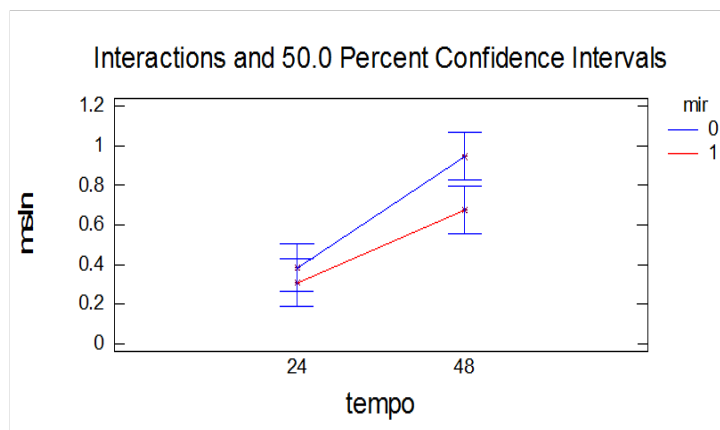


Figura J.7: Interaction plot tra i valori di espressione della mesotelina in funzione tempo e del trattamento per l'esperimento 4. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: tempo	0,64635	1	0,64635	7,44	0,02600
B: mir	0,08776	1	0,08776	1,01	0,3444
Interactions					
AB	0,02865	1	0,02865	0,33	0,5817
RESIDUAL	0,69540	8	0,08692		
TOTAL (corrected)	1,45816	11			

Tabella J.13: Analisi della varianza per i valori della mesotelina registrati per l'esperimento 4

Il quarto ed il quinto esperimento sono stati effettuati per soli due tempi sperimentali, 24 e 48 ore dopo la trafezione, visto che dalle prove precedenti il segnale più forte si aveva in corrispondenza delle 48 ore, mentre alle 24 e 72 ore il trattamento con il microRNA non portava a grandi differenze nei livelli del messaggero. Si sono scelte le 24 ore, e gli esperimenti hanno coinvolto tre repliche per condizione sperimentale. Si può osservare come alle 48h ci sia una significativa differenza tra le cellule trattate con il microRNA e quelle di controllo, laddove alle 24h i valori registrati sono circa i medesimi.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	12	0,57779			
tempo					
24	6	0,34571	0,12036	0,26068	0,43073
48	6	0,80987	0,12036	0,72485	0,89490
mir					
0	6	0,66331	0,12036	0,57829	0,74833
1	6	0,49227	0,12036	0,40725	0,57730
tempo by mir					
24; 0	3	0,38237	0,17022	0,26213	0,50261
24; 1	3	0,30905	0,17022	0,18881	0,42929
48; 0	3	0,94425	0,17022	0,82401	1,06449
48; 1	3	0,67550	0,17022	0,55525	0,79574

Tabella J.14: Tabella dei least squares means per la mesotelina nell'esperimento 4, con intervalli di confidenza del 50%

J.2.5 ANOVA 2 fattori, Esperimento 5

Interaction Plot

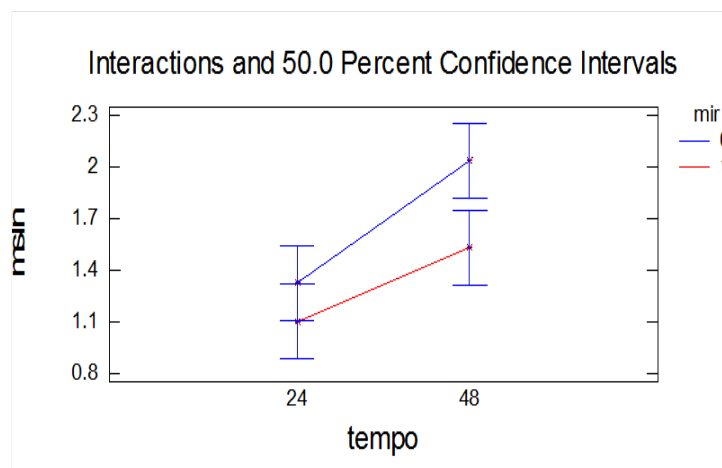


Figura J.8: Interaction plot tra i valori di espressione della mesotelina in funzione tempo e del trattamento per l'esperimento 5. Mir0= controlli negativi; Mir1= trattamento con miR204-5p

Analisi della varianza

É questo l'unico esperimento in cui la variabile sperimentale "tempo" non risulta signi-

Analysis of Variance for msln					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main Effects					
A: tempo	0,97444	1	0,97444	3,47	0,09970
B: mir	0,40057	1	0,40057	1,42	0,26680
Interactions					
AB	0,05977	1	0,05977	0,21	0,65700
RESIDUAL	2,24917	8	0,28115		
TOTAL (corrected)	3,68395	11			

Tabella J.15: Analisi della varianza per i valori della mesotelina registrati per l'esperimento 5

ficativa, ma è cmq evidente il trend cinetico di aumento di espressione del gene alle 48h, e dell'effetto del trattamento, per quanto non venga raggiunta la significatività statistica.

Table of Least Squares Means for msln					
Level	Count	Mean	Std Error	Lower limit	Upper limit
Grand Mean	12	1,49908			
tempo					
24	6	1,21412	0,21647	1,06121	1,36703
48	6	1,78404	0,21647	1,63113	1,93695
mir					
0	6	1,68179	0,21647	1,52888	1,83470
1	6	1,31638	0,21647	1,16347	1,46929
tempo by mir					
24; 0	3	1,32625	0,30613	1,11001	1,54250
24; 1	3	1,10199	0,30613	0,88574	1,31824
48; 0	3	2,03732	0,30613	1,82107	2,25357
48; 1	3	1,53077	0,30613	1,31452	1,74701

Tabella J.16: Tabella dei least squares means per la mesotelina nell'esperimento 5, con intervalli di confidenza del 50%

Bibliografia

- Adamson I. Y.; Bakowska J.; Bowden D. H. (1993). Mesothelial cell proliferation after instillation of long or short asbestos fibers into mouse lung. *The American journal of pathology*, **142**(4), 1209–16.
- Altuvia Y.; Landgraf P.; Lithwick G.; Elefant N.; Pfeffer S.; Aravin A.; Brownstein M. J.; Tuschl T.; Margalit H. (2005). Clustering and conservation patterns of human microRNAs. *Nucleic acids research*, **33**(8), 2697–2706.
- Bagga S.; Bracht J.; Hunter S.; Massirer K.; Holtz J.; Eachus R.; Pasquinelli A. E. (2005). Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation. *Cell*, **122**(4), 553–563.
- Bartel D. P.; Lee R.; Feinbaum R. (2004). MicroRNAs : Genomics , Biogenesis , Mechanism , and Function Genomics. *Cell*, **116**, 281–297.
- Basyuk E.; Suavet F.; Doglio A.; Bordonné R.; Bertrand E. (2003). Human let-7 stem-loop precursors harbor features of RNase III cleavage products. *Nucleic acids research*, **31**(22), 6593–6597.
- Behm-Ansmant I.; Rehwinkel J.; Doerks T.; Stark A.; Bork P.; Izaurralde E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4: NOT deadenylase and DCP1: DCP2 decapping complexes. *Genes & development*, **20**(14), 1885–1898.
- Bera T. K.; Pastan I. (2000). Mesothelin is not required for normal mouse development or reproduction. *Molecular and cellular biology*, **20**(8), 2902–6.
- Bharadwaj U.; Li M.; Chen C.; Yao Q. (2008). Mesothelin-induced pancreatic cancer cell proliferation involves alteration of cyclin E via activation of signal transducer and activator of transcription protein 3. *Molecular Cancer Research*, **6**(11), 1755–1765.

- Carroll S. B.; Prud'homme B.; Gompel N. (2008). Regulating evolution. *Scientific American*, **298**(5), 60–7.
- Chang K.; Pastan I. (1996). Molecular cloning of mesothelin, a differentiation antigen present on mesothelium, mesotheliomas, and ovarian cancers. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(1), 136–40.
- Chen K.; Rajewsky N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature reviews. Genetics*, **8**(2), 93–103.
- Chen S.-H.; Hung W.-C.; Wang P.; Paul C.; Konstantopoulos K. (2013). Mesothelin binding to CA125/MUC16 promotes pancreatic cancer cell motility and invasion via MMP-7 activation. *Scientific reports*, **3**, 1870.
- Chiang H. R.; Schoenfeld L. W.; Ruby J. G.; Auyeung V. C.; Spies N.; Baek D.; Johnston W. K.; Russ C.; Luo S.; Babiarz J. E.; Belloch R.; Schroth G. P.; Nusbaum C.; Bartel D. P. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & Development*, **24**(10), 992–1009.
- Coller J.; Parker R. (2004). Eukaryotic mRNA decapping. *Annual review of biochemistry*, **73**(1), 861–890.
- Daniels R. J.; Peden J. F.; Lloyd C.; Horsley S. W.; Clark K.; Tufarelli C.; Kearney L.; Buckle V. J.; Doggett N. a.; Flint J.; Higgs D. R. (2001). Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16. *Human molecular genetics*, **10**(4), 339–52.
- Diederichs S.; Haber D. A. (2007). Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*, **131**(6), 1097–1108.
- Dunham I.; Hunt A.; Collins J. (1999). The DNA sequence of human chromosome 22. *Nature*, **402**(December), 489–496.
- Emes R. D. (2003). Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Human Molecular Genetics*, **12**(7), 701–709.
- Enerly E.; Steinfeld I.; Kleivi K.; Leivonen S.-K.; Aure M. R.; Russnes H. G.; Rønneberg J. A.; Johnsen H.; Navon R.; Rødland E.; Others (2011). miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PloS one*, **6**(2), e16915.
- Eulalio A.; Rehwinkel J.; Stricker M.; Huntzinger E.; Yang S.-F.; Doerks T.; Dörner S.; Bork P.; Boutros M.; Izaurralde E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes & development*, **21**(20), 2558–2570.

- Eulalio A.; Huntzinger E.; Nishihara T. (2009). Deadenylation is a widespread effect of miRNA regulation. *Rna*, pp. 21–32.
- Fabian M. R.; Mathonnet G.; Sundermeier T.; Mathys H.; Zipprich J. T.; Svitkin Y. V.; Rivas F.; Jinek M.; Wohlschlegel J.; Doudna J. A.; Others (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Molecular cell*, **35**(6), 868–880.
- Fabian M. R.; Sonenberg N.; Filipowicz W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry*, **79**, 351–79.
- Fang Z.; Rajewsky N. (2011). The impact of miRNA target sites in coding sequences and in 3'UTRs. *PloS one*, **6**(3), e18067.
- Gilad Y.; Man O.; Pääbo S.; Lancet D. (2003). Human specific loss of olfactory receptor genes. *Proceedings of the National Academy of Sciences*, **100**(6), 3324–3327.
- Gottardo F.; Liu C. G.; Ferracin M.; Calin G. A.; Fassin M.; Bassi P.; Sevignani C.; Byrne D.; Negrini M.; Pagano F.; Others (2007). Micro-RNA profiling in kidney and bladder cancers. *Urologic Oncology: Seminars and Original Investigations*, **25**(5), 387–392.
- Grimson A.; Farh K. K.-H.; Johnston W. K.; Garrett-Engele P.; Lim L. P.; Bartel D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, **27**(1), 91–105.
- Guang S.; Bochner A. F.; Pavelec D. M.; Burkhart K. B.; Harding S.; Lachowiec J.; Kennedy S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science*, **321**(5888), 537–541.
- Gubbels J. a. a.; Belisle J.; Onda M.; Rancourt C.; Migneault M.; Ho M.; Bera T. K.; Connor J.; Sathyanarayana B. K.; Lee B.; Pastan I.; Patankar M. S. (2006). Mesothelin-MUC16 binding is a high affinity, N-glycan dependent interaction that facilitates peritoneal metastasis of ovarian tumors. *Molecular cancer*, **5**(1), 50.
- Guled M.; Lahti L.; Lindholm P. M.; Salmenkivi K.; Bagwan I.; Nicholson A. G.; Knuutila S. (2009). CDKN2A, NF2, and JUN are dysregulated among other genes by miRNAs in malignant mesothelioma—A miRNA microarray analysis. *Genes, Chromosomes and Cancer*, **48**(7), 615–623.
- Hahn Y.; Jeong S.; Lee B. (2007). Inactivation of MOXD2 and S100A15A by exon deletion during human evolution. *Molecular biology and evolution*, **24**(10), 2203–2212.

- Han J.; Lee Y.; Yeom K.-H.; Kim Y.-K.; Jin H.; Kim V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes & development*, **18**(24), 3016–3027.
- Hassan R.; Bera T.; Pastan I. (2004). Mesothelin : A New Target for Immunotherapy. *Clinical Cancer Research*, **10**, 3937–3942.
- Havens M. a.; Reich A. a.; Duelli D. M.; Hastings M. L. (2012). Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic acids research*, **40**(10), 4626–40.
- Hobolth A.; Dutheil J. Y.; Hawks J.; Schierup M. H.; Mailund T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome research*, **21**(3), 349–56.
- Hucl T.; Brody J.; Gallmeier E. (2007). High Cancer-Specific Expression of Mesothelin (MSLN) Is Attributable to an Upstream Enhancer Containing a Transcription Enhancer Factor-Dependent MCAT Motif. *Cancer research*, **67**(19), 9055–9065.
- Jongsma J.; van Montfort E.; Vooijs M.; Zevenhoven J.; Krimpenfort P.; van der Valk M.; van de Vijver M.; Berns A. (2008). A conditional mouse model for malignant mesothelioma. *Cancer cell*, **13**(3), 261–71.
- Jukes T. H.; Kimura M. (1984). Evolutionary constraints and the neutral theory. *Journal of molecular evolution*, **21**(1), 90–2.
- Kaneko O.; Gong L.; Zhang J.; Hansen J. K.; Hassan R.; Lee B.; Ho M. (2009). A binding domain on mesothelin for CA125/MUC16. *The Journal of biological chemistry*, **284**(6), 3739–49.
- Kim D.; Hahn Y. (2011). Exon deletion in the MSLN gene encoding MPF/mesothelin precursor protein during Laurasiatherian mammal evolution. *International journal of molecular medicine*, **27**(2), 289–293.
- Kojima T.; Oh-eda M.; Hattori K.; Taniguchi Y.; Tamura M.; Ochi N.; Yamaguchi N. (1995). Molecular Cloning and Expression of Megakaryocyte Potentiating Factor cDNA. *Journal of Biological Chemistry*, **270**(37), 21984–21990.
- Lagos-Quintana M.; Rauhut R.; Lendeckel W.; Tuschl T. (2001). Identification of novel genes coding for small expressed RNAs. *Science*, **294**(5543), 853–858.
- Lee C.-T.; Risom T.; Strauss W. M. (2007). Evolutionary Conservation of microRNA Regulation Circuits: An examination of microRNA gene complexity and conserved microRNA-target interactions through Metazoan phylogeny. *DNA and Cell Biology*, **26**(4), 209–218.

- Lee Y.; Ahn C.; Han J.; Choi H.; Kim J.; Yim J.; Lee J.; Provost P.; Rasmussen O.; Kim S.; Others (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**(6956), 415–419.
- Lee Y.; Kim M.; Han J.; Yeom K.-H.; Lee S.; Baek S. H.; Kim V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, **23**(20), 4051–60.
- Lim L. P.; Lau N. C.; Garrett-Engle P.; Grimson A.; Schelter J. M.; Castle J.; Bartel D. P.; Linsley P. S.; Johnson J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**(7027), 769–773.
- Lund E.; Güttinger S.; Calado A.; Dahlberg J. E.; Kutay U. (2004). Nuclear export of microRNA precursors. *Science*, **303**(5654), 95–98.
- Marsella J. M.; Liu B. L.; Vaslet C. a.; Kane a. B. (1997). Susceptibility of p53-deficient mice to induction of mesothelioma by crocidolite asbestos fibers. *Environmental health perspectives*, **105 Suppl**(September), 1069–72.
- Matsui A.; Go Y.; Niimura Y. (2010). Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. *Molecular biology and evolution*, **27**(5), 1192–1200.
- Melaiu O.; Cristaudo A.; Melissari E.; Di Russo M.; Bonotti A.; Bruno R.; Foddiss R.; Gemignani F.; Pellegrini S.; Landi S. (2011). A review of transcriptome studies combined with data mining reveals novel potential markers of malignant pleural mesothelioma. ... *Research/Reviews in ...*, **750**(2), 132–140.
- Ming-Cheng C.; Chi-An C. (2012). Mesothelin enhances invasion of ovarian cancer by inducing MMP-7 through MAPK/ERK and JNK pathways. *Biochemical ...*, **442**(2), 293–302.
- Mishima Y.; Giraldez A. J.; Takeda Y.; Fujiwara T.; Sakamoto H.; Schier A. F.; Inoue K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Current biology*, **16**(21), 2135–2142.
- Nelson H.; Almquist L.; LaRocca J. (2011). The relationship between tumor MSLN methylation and serum mesothelin (SMRP) in mesothelioma. *Epigenetics*, **6**(8), 1029–34.
- Nobrega M. a.; Pennacchio L. a. (2004). Comparative genomic analysis as a tool for biological discovery. *The Journal of physiology*, **554**(Pt 1), 31–9.

- Nottrott S.; Simard M. J.; Richter J. D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nature structural & molecular biology*, **13**(12), 1108–1114.
- Olsen P. H.; Ambros V. (1999). The lin-4 Regulatory RNA Controls Developmental Timing in *Caenorhabditis elegans* by Blocking LIN-14 Protein Synthesis after the Initiation of Translation. *Developmental biology*, **216**(2), 671–680.
- Orom U. A.; Nielsen F. C.; Lund A. H. (2008). MicroRNA-10a binds the 5' UTR of ribosomal protein mRNAs and enhances their translation. *Molecular cell*, **30**(4), 460–471.
- Petersen C. P.; Bordeleau M.-E.; Pelletier J.; Sharp P. A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Molecular cell*, **21**(4), 533–542.
- Place R. F.; Li L.-C.; Pookot D.; Noonan E. J.; Dahiya R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences*, **105**(5), 1608–1613.
- Rhoades M. W.; Reinhart B. J.; Lim L. P.; Burge C. B.; Bartel B.; Bartel D. P. (2002). Prediction of plant microRNA targets. *Cell*, **110**(4), 513–520.
- Robinson C.; van Bruggen I.; Segal A.; Dunham M.; Sherwood A.; Koentgen F.; Robinson B. W. S.; Lake R. a. (2006). A novel SV40 Tag transgenic model of asbestos-induced mesothelioma: malignant transformation is dose dependent. *Cancer research*, **66**(22), 10786–94.
- Rodriguez A.; Griffiths-Jones S. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome ...*, pp. 1902–1910.
- Rump A.; Morikawa Y.; Tanaka M.; Minami S.; Umesaki N.; Takeuchi M.; Miyajima A. (2004). Binding of ovarian cancer antigen CA125/MUC16 to mesothelin mediates cell adhesion. *The Journal of biological chemistry*, **279**(10), 9190–8.
- Russo V. E. A.; Martienssen R. A.; Riggs A. D.; Others (1996). *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press.
- Salzman D. W.; Shubert-Coleman J.; Furneaux H. (2007). P68 RNA helicase unwinds the human let-7 microRNA precursor duplex and is required for let-7-directed silencing of gene expression. *Journal of Biological Chemistry*, **282**(45), 32773–32779.
- Sandhu H.; Dehnen W.; Roller M.; Abel J.; Unfried K. (2000). mRNA expression patterns in different stages of asbestos-induced carcinogenesis in rats. *Carcinogenesis*, **21**(5), 1023–1029.

- Sathyanarayana B. K.; Hahn Y.; Patankar M. S.; Pastan I.; Lee B. (2009). Mesothelin, Stereocilin, and Otoancorin are predicted to have superhelical structures with ARM-type repeats. *BMC structural biology*, **9**, 1.
- Scholler N.; Garvik B.; Hayden-Ledbetter M.; Kline T.; Urban N. (2007). Development of a CA125-mesothelin cell adhesion assay as a screening tool for biologics discovery. *Cancer letters*, **247**(1), 130–6.
- Schwarz D. S.; Hutvagner G.; Du T.; Xu Z.; Aronin N.; Zamore P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**(2), 199–208.
- Selbach M.; Schwanhäusser B.; Thierfelder N.; Fang Z.; Khanin R.; Rajewsky N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**(7209), 58–63.
- Servais E. L.; Colovos C.; Rodriguez L.; Bograd A. J.; Nitadori J.-i.; Sima C.; Rusch V. W.; Sadelain M.; Adusumilli P. S. (2012). Mesothelin overexpression promotes mesothelioma cell invasion and MMP-9 secretion in an orthotopic mouse model and in epithelioid pleural mesothelioma patients. *Clinical Cancer Research*, **18**(9), 2478–2489.
- Tan K.; Kajino K.; Momose S.; Masaoka A.; Sasahara K.; Shiomi K.; Izumi H.; Abe M.; Ohtsui N.; Wang T.; Hino O.; Fujii H. (2010). Mesothelin (MSLN) promoter is hypomethylated in malignant mesothelioma, but its expression is not associated with methylation status of the promoter. *Human pathology*, **41**(9), 1330–8.
- Tavaré S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci*, **17**, 57–86.
- Tay Y.; Zhang J.; Thomson A. M.; Lim B.; Rigoutsos I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**(7216), 1124–1128.
- Tycowski K. T.; Shu M.-D.; Steitz J. A. (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & development*, **7**(7a), 1176–1190.
- Uddalak B.; Christian M.; Min L.; Changyi C.; Qizhi Y. (2011). Mesothelin confers pancreatic cancer cell resistance to TNF- α -induced apoptosis through Akt/PI3K/NF- κ B activation and IL-6/Mcl-1 overexpression. *Molecular Cancer*, **10**(1), 106.
- Urwin D.; Lake R. (2000). Structure of the Mesothelin/MPF gene and characterization of its promoter. *Molecular Cell Biology Research Communications*, **3**(1), 26–32.

- Wang K.; Bodempudi V.; Liu Z. (2012a). Inhibition of mesothelin as a novel strategy for targeting cancer cells. *PLoS One*, **7**(4), e33214.
- Wang Y.; Wang L.; Li D. (2012b). Mesothelin Promotes Invasion and Metastasis in Breast Cancer Cells. *Journal of International ...*, pp. 2109–2116.
- Whelan S.; Goldman N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution*, **18**(5), 691–9.
- Wightman B.; Ha I.; Ruvkun G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, **75**(5), 855–862.
- Wu L.; Belasco J. G. (2005). Micro-RNA regulation of the mammalian *lin-28* gene during neuronal differentiation of embryonal carcinoma cells. *Molecular and cellular biology*, **25**(21), 9198–9208.
- Wu L.; Fan J.; Belasco J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(11), 4034–4039.
- Yamaguchi N.; Hattori K.; Oh-eda M. (1994). A novel cytokine exhibiting megakaryocyte potentiating activity from a human pancreatic tumor cell line HPC-Y5. *Journal of Biological ...*, pp. 805–808.
- Yi R.; Qin Y.; Macara I. G.; Cullen B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development*, **17**(24), 3011–3016.
- Zhang D.; Kobayashi T.; Kojima T.; Kanenishi K.; Hagiwara Y.; Abe M.; Okura H.; Hamano Y.; Sun G.; Maeda M.; Jishage K.-i.; Noda T.; Hino O. (2011). Deficiency of the *Erc/mesothelin* gene ameliorates renal carcinogenesis in *Tsc2* knockout mice. *Cancer science*, **102**(4), 720–7.